

中科天玑
数据能力介绍





总体能力

境内外开源数据采集能力

采集分类	采集通道	信源数量	采集信息	采集规模
互联网	新闻	1.5万+网站	文章：标题、内容、发布时间、URL、网站、作者等 评论：用户昵称、评论内容、评论时间、头像、原文id、原文URL等	200万条/日
	电子报	1000+网站	文章：标题、内容、发布时间、URL、网站、作者等	2万+/日
	智库	150+中外知名智库	帖子：标题、内容、发布时间、URL、网站、作者、报告附件地址等	1000+/日
两微一端	微博	2亿+用户	微博信息：微博ID、微博内容、发布时间、用户ID、用户昵称、用户头像、微博URL、转发数、评论数、点赞数、图片列表等	7000万+/日
	微信	200万+活跃用户	微信公众号文章：标题、内容、发布时间、URL、公众号id、公众号名称等	180万+/日
	APP移动端	100+APP	文章：标题、内容、发布时间、URL、网站、作者等 评论：用户昵称、评论内容、评论时间、头像、原文id、原文URL等	300万/日
短视频	抖音、快手、火山、西瓜、好看、美拍、bilibili	2亿+用户	视频信息（发布时间、视频url、用户ID、用户名、关注数、粉丝数、视频ID、转发数、收藏数、弹幕数、视频分享数、观看数、播放量、评论数量、视频封面图链接、视频简介、无水印下载链接。）、用户信息	1.2亿+/日
互动社区	论坛	8万+板块	帖子信息：标题、内容、发布时间、URL、网站、作者等	30万+/日
	问答	2千+板块	帖子：标题、内容、发布时间、URL、网站、作者等	1万+/日
JW	境外新闻	1万+网站	文章：标题、内容、发布时间、URL、网站、作者等	15万/日
	Twitter	5千万+用户	推文信息：推文ID、推文内容、推文创建时间、推文是否原创、推文转发数、推文喜爱数、用户ID、用户名称、用户头像、用户注册时间、用户粉丝数、用户关注数、用户推文数、用户喜爱数等 个人基本信息：人物名、人物屏幕名称、人物粉丝数目、人物好友数目、人物简介、人物创建时间、人物头像、人物位置信息、人物帖子数目等; 好友或粉丝信息：人物名、人物屏幕名称、人物粉丝数目、人物好友数目、人物简介、人物创建时间、人物头像、人物位置信息、人物帖子数目等 评论转推信息：推文ID、推文评论内容、推文创建时间、转推人用户名称、转推人用户头像、转推人注册时间、转推用户粉丝数、转推人关注数、转推人推文数、转推人喜爱数等	400万+/日
	Facebook	10万+用户	推文信息：推文ID、推文内容、推文创建时间、推文图片URL列表、推文评论数、推文转发数等 个人基本信息：人物名、人物屏幕名称、人物头像、个人描述、个人主页地址等 好友信息：人物名、人物屏幕名称、人物头像、个人描述、个人主页地址等	10万+/日
	Instagram	1000+主流媒体和大V账号	帖子信息：推文ID、推文内容、推文创建时间、推文图片URL列表、推文评论数、推文转发数等	4万+/日
	Telegram	1000+主流媒体和大V账号	帖子信息：推文ID、推文内容、推文创建时间、推文图片URL列表等	8万+/日
	YouTube	1万+账号	视频信息：视频ID、视频封面图链接、视频发布时间、视频描述、播放量、评论数、点赞数、视频地址等	
	领英	4亿+账号	用户履历信息：名字、头像、描述、个人技能、教育经理、工作经历等	——
	mastodon	——	推文信息：推文ID、推文内容、推文创建时间、推文是否原创、推文转发数、推文喜爱数、用户ID、用户名称、用户头像、用户注册时间、用户粉丝数、用户关注数、用户推文数、用户喜爱数等 个人基本信息：人物名、人物屏幕名称、人物粉丝数目、人物好友数目、人物简介、人物创建时间、人物头像、人物帖子数目等;	10万+/日

M和T方向的网站2000个，账号8W+，
重点账号5000个

重点账号积累情况

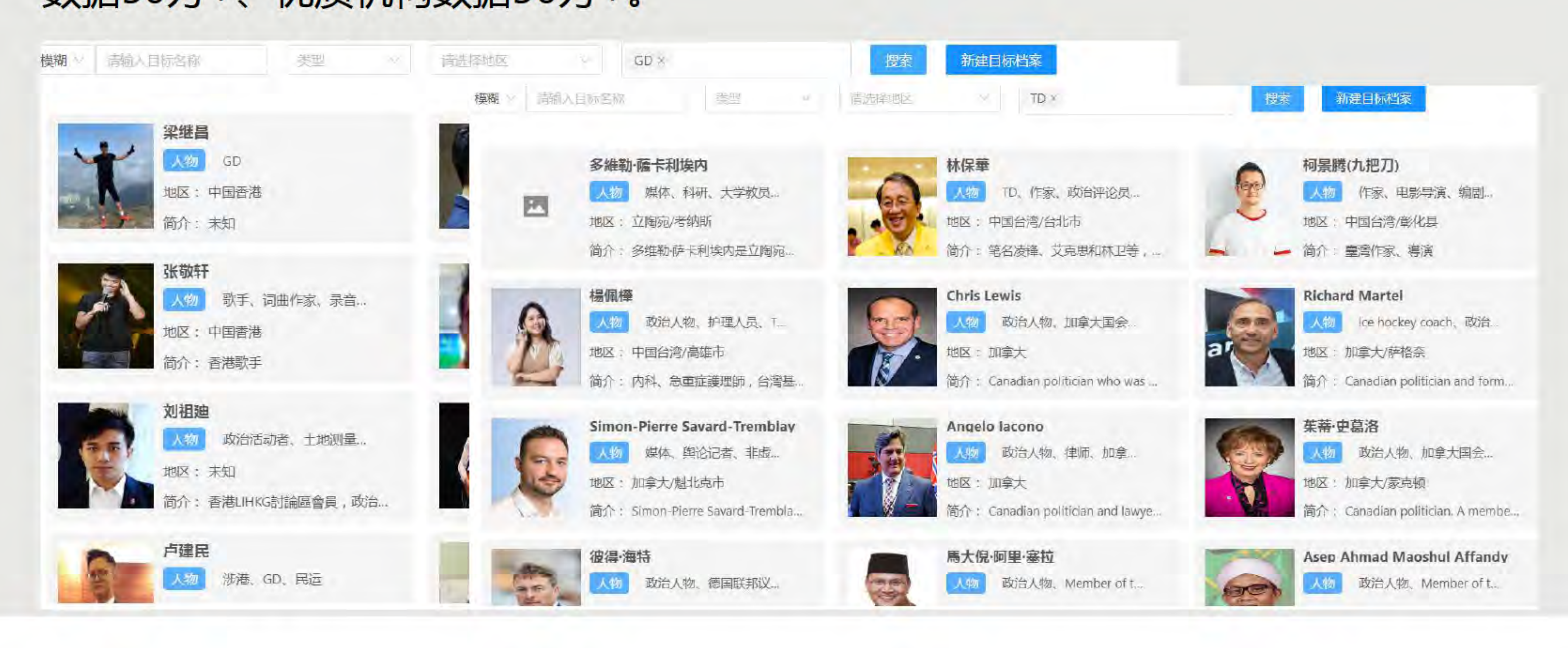
国内重点账号2000+。FH NGO近千个，FH个人账号近千个，FH组织及其成员账号数百个：账号、姓名、性别、学历、住址、FH活动等。涉K、涉邪等账号数百个。

序号	姓名（必填）	外文名	所在机构	职务	性别	出生日期	国家	省/州	市/区	个人简介
1	黎智英	Jimmy Lai Chee-ying	壹传媒创始人	香港壹传媒创始人	男	1948/12/8	中国	香港		香港企业家，佐
2	陈方安生	Chan Fang On Sang		首位华人布政司	女	1940/1/17	中国	香港		陈方安生曾经是
3	戴耀廷	Benny Tai Yiu-ting	香港大学	香港大学副教授	男	1964/7/12	中国	香港		戴耀廷中学毕业
4	李柱铭	Martin Lee Chu-ming	香港公民党	香港民主党创党主席	男	1938/6/8	中国	香港		广东惠州人，成
5	何俊仁	Albert Ho Chun-Yan	香港公民党	香港民主党前主席|香港立法会议员	男	1951/12/1	中国	香港		广东中山人，成
6	李永达	Lee Wing-tat	香港公民党	香港民主党前主席|香港立法会议员|葵青	男	1955/12/25	中国	香港		广东惠阳人，出
7	胡志伟	Wu Chi-wai	香港公民党	香港民主党主席|香港民主党立法会议员	男	1962/10/18	中国	香港		香港出生，前黄
8	林卓廷	Lam Cheuk Ting	公民党立法会	香港民主党立法会议员	男	1977/6/13	中国	香港	屯门	香港立法会议员、
9	许智峰	Hui Chi-fung	公民党立法会	香港民主党立法会议员	男	1982/6/8	中国	香港	屯门	现任香港中西区
10	尹兆坚	Andrew Wan	公民党立法会	香港民主党立法会议员	男	1969/6/7	中国	香港	九龙	香港政治人物，
11	涂谨申	James To Kun-sun	公民党立法会	香港民主党立法会议员	男	1963/3/11	中国	香港	九龙	香港大学法律系
12	黄碧云	Helena Wong Pik-wan	公民党立法会	香港民主党立法会议员	女	1959/3/21	中国	香港	九龙	现任香港九龙西
13	邝俊宇	Roy Kwong Chun-yu	香港公民党	香港民主党中央委员会委员	男	1983/2/9	中国	香港	元朗	香港立法会议员、
14	梁家杰	Alan Leong Kah-kit	香港公民党	香港公民党主席|香港立法会议员	男	1958/2/22	中国	香港	旺角	广东新会人，香
15	杨岳桥	Alvin Yeung	公民党立法会	香港公民党立法会议员	男	1981/6/5	中国	香港	元朗	香港执业大律师，
16	郭荣铿	Dennis Kwok Wing-hang	公民党立法会	香港公民党立法会议员	男	1978/4/15	中国	香港		香港政治人物、
17	毛孟静	Claudia Mo Man-ching	公民党立法会	香港公民党立法会议员	女	1957/1/18	中国	香港	跑马地	祖籍浙江宁波，教、家庭教会
18	谭文豪	Jeremy Tam Man-ho	公民党立法会	香港公民党立法会议员	男	1975/6/13	中国	香港		出生于香港，祖
19	陈淑庄	Tanya Chan	公民党立法会	香港公民党立法会议员	女	1971/9/14	中国	香港		祖籍上海，香港
20	郭永健	Steven Kwok Wing-kin	香港工党	香港工党主席	男	1986/11/30	中国	香港		祖籍广东揭阳，
21	李卓人	Lee Cheuk-yan	香港工党	香港工党副主席	男	1957/2/12	中国	香港		广东潮阳人，上

			纽约时报》专栏作者	
邓聿文	@dyw1968316	https://twitter.com/dyw1968316	政治评论员、海外中文媒体撰稿人	中国政府、中国对外关系
滕彪	@tengbiao	https://twitter.com/tengbiao	维权律师，"零八宪章"发起人之一，曾响应所谓"茉莉花革命"	人权、死刑、中国政府

虚实映射能力

虚实映射的人物组织数据库，基于wiki结构化生成，人工打标补充完善，优质人物数据50万+、优质机构数据50万+。





对M数据

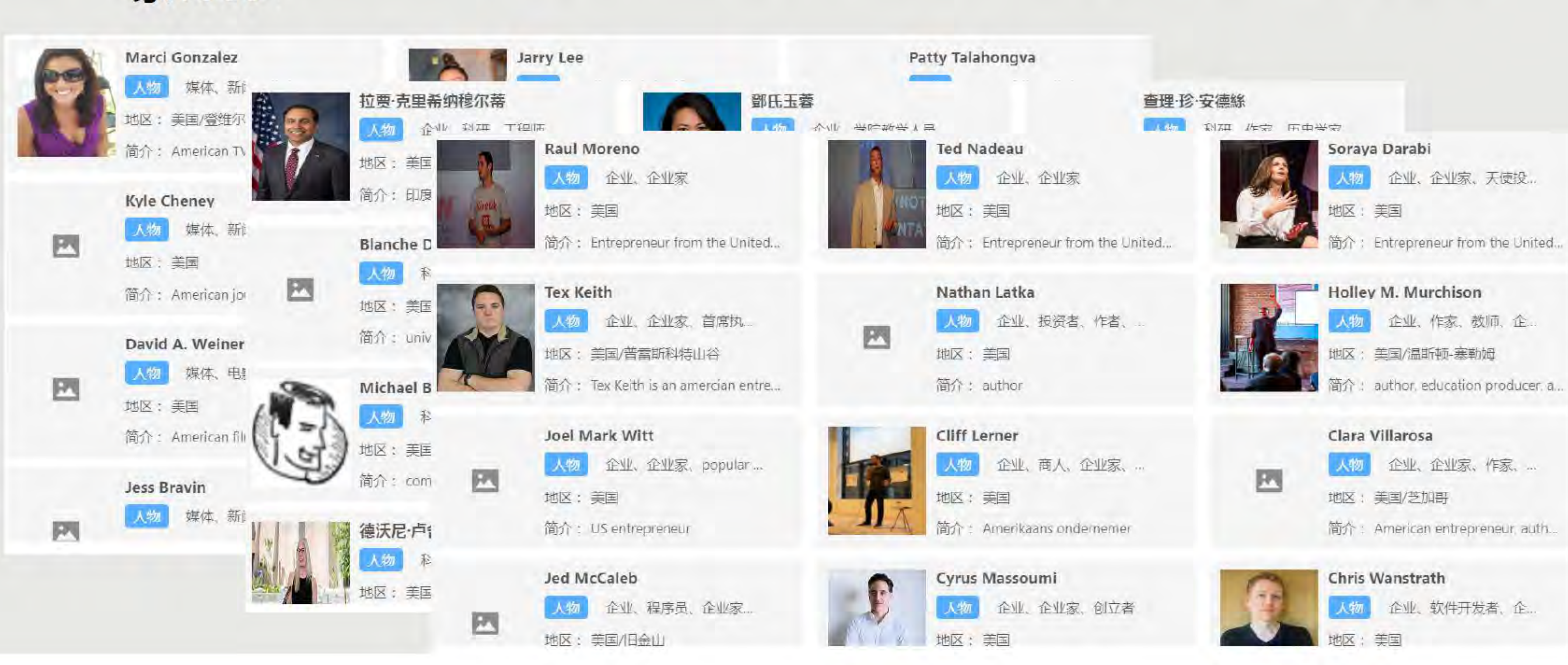
涉M数据积累：政要数据

117届国会议员、2022年国会候选人、建制派、特朗普支持者、右翼势力4000+、NGO账号数据数百个、其他A政治人物和意见领袖2000+

类别	党籍	两院	职位	英文名-去空格	英文名	author_name
右翼势力	共和党			Abraham Hamade	Abraham Hamadeh	Abe Hamadeh for Arizona AG
右翼势力	共和党			Adam Laxalt	Adam Laxalt	Adam Paul Laxalt
右翼势力	共和党			Alison Hayden	Alison Hayden	Alison Hayden 賀亞珊us CA-14 vs.Eric Swalwell
右翼势力	共和党			Andy Biggs	Andy Biggs	Andy Biggs
右翼势力	共和党			Barry Loudermilk R	Barry Loudermilk R	Barry Loudermilk
右翼势力	共和党			Ben Cline	Ben Cline	Ben Cline
右翼势力	共和党			Bill Posey	Bill Posey	Bill Posey
右翼势力	共和党			Brian Babin	Brian Babin	Brian Babin
右翼势力	共和党			Brian Perras	Brian Perras	Brian Perras
右翼势力	共和党			Byron Donalds	Byron Donalds	Byron Donalds
右翼势力	共和党			Chip Roy	Chip Roy	Chip Roy
右翼势力	共和党			Bob Good	Bob Good	Congressman Bob Good
右翼势力	共和党			Dan Bishop	Dan Bishop	Dan Bishop
右翼势力	共和党			Dan Cox	Dan Cox	Dan Cox us - Delegate & Candidate for Governor
右翼势力	共和党			Darren Bailey	Darren Bailey	Darren Bailey for Governor
右翼势力	共和党			Darren Aquino	Darren Aquino	Darren.Dione.Aquino usCONGRESSusUS.House Elec
右翼势力	共和党			Debbie Lesko	Debbie Lesko	Debbie Lesko
右翼势力	共和党			Gary Palmer	Gary Palmer	Gary Palmer
右翼势力	共和党			Greg Steube	Greg Steube	Greg Steube
右翼势力	共和党			Jim Marchant	Jim Marchant	Jim Marchant

涉M数据积累：专业人物账号

新闻工作者5000+、专家学者1000+、经济学家400+、科研人员1W+、企业家1000+



涉M数据积累：专业机构账号

科研机构8000+、军事机构200+、企业2W+



德克萨斯基督教大学
科研机构　科研、私立大学、...
地区：美国
简介：private university in Fort Wort...



普林斯顿大学工程与应用科学学院
科研机构　科研、科研机构、...



耶鲁大学工程与应用科学学院
科研机构　科研、研究所

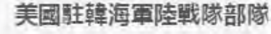
斯坦福大学工程学院
科研机构　科研、研究所
地区：美国
简介：Welcome to the Stanford Sch...



美國陸軍電子競技
军事机构　电玩游戏战队、军...
地区：美国
简介：esports team sponsored by th...

VR-55
军事机构　军事单位
地区：美国
简介：military unit

美國駐韓海軍陸戰隊部隊
军事机构　军事、司令部
地区：美国



简介：미국 해병대의 한반도 전구 사...



加利福尼亞大學聖巴巴拉分校
科研机构　科研、大学、美国...
地区：美国
简介：1891年創建的一所學校



United States Forces Afghanis...
军事机构　一体化作战司令部
地区：美国
简介：미국 중부사령부 예하 통합전...



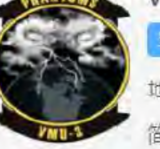
3d Space Experimentation Sq...
军事机构　军事单位
地区：美国
简介：military unit



第132聯隊
军事机构　军事单位
地区：美国
简介：military unit

Boston University School of ...
科研机构　学院





VMU-3
军事机构　军事单位、飞行中...
地区：美国
简介：unmanned aerial vehicle squa...





United States Naval Sea Cadet...
军事机构　军事单位
地区：美国
简介：U.S. Navy sponsored organiza...



坦克汽车研究、开发和工程中心
军事机构　军事单位
地区：美国
简介：military unit

軍眷與士氣福利休閒計畫
军事机构　军事单位
地区：美国



西南師
军事机构　军事单位
地区：美国

南大西洋師
军事机构　军事单位
地区：美国



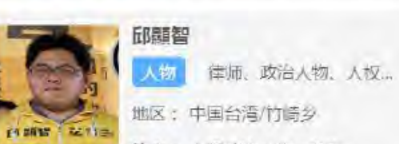
对T数据

涉T数据积累：政要数据

T政要数据，含党派、电话、传真、邮箱、出生地点、工作地点、信仰、教育经历、社会关系等。

name_cn	name_en	name_alia	nickname	标签	professior	wiki_cn	wiki_en	potrait	country	party_nam	organ	post	birth_plac	birth_date	residence	sex	brief_intrc	pul
林佳龙	Lin Chia-lu		龙哥、龙龙	政客	政治学	https://zh.w	https://ww	\10.jpg	中国台湾	民主进步党			台北市万华	1964-02-1	台中市西区	男	林佳龙（台	《;
卓伯源				政客		https://zh.w		\1050.jpg	中国台湾	中国国民党			彰化县田中	1965-03-2		男	卓伯源（1	
张亚中				政客、国际		https://zh.w		\1051.jpg	中国台湾	中国国民党	孙文学校	总校长	台中市	1954-12-2		男	张亚中（1	
陈时中	Chen Shih-		阿中部长	政客、牙医		http												
黄伟哲	Huang Wei		阿哲市长	政客		http												
陈其迈	Chen Chi-n		迈迈、暖男	政客		http												
林智坚	Lin Chih-ch			政客		http												
赖清德	Lai Ching-t			政客、内科		http												
蔡英文		Tsai Ing-w	小英、辣台	政客、大学	竞争法、国	http												
李大维				政客、外交		http												
苏贞昌			电火球、不	政客、前律		http												
沈荣津				政客	工业管理	http												
李孟谚				政客		http												
游锡堃			水牛伯、台	政客		http												
蔡其昌				政客		http												
林志嘉				政客		http												
许宗力				政客、著名		http												
周弘宪				政客		http												
刘建忻				政客		http												
陈菊			花妈、菊姐	政客		http												
朱富美				政客														




涉T知识图谱

T政府人物组织图谱，包括组织名称、成立时间、创始人、负责人、政治倾向、成员规模、关注领域、电话、地址、职能、对华态度、坐标；人物姓名、婚姻状况、出生日期、性别、党派、政治倾向、受教育程度、民族、职务、工作单位、教育经历等。

组织名称	简介(字符串)	成立时间	创始人	负责人(与	政治倾向	成员规模	关注领域	电话	邮箱	地址	组织架构
新党	新党是中华民国的一	1993年8月10日	赵少康、郁慕	吴成典	蓝营	约500人	政治、民主	(02)2756-2222		台北市松山区光复南路	郁慕明曾经把新党
民主进步党	民主进步党（英语:	1986年9月28日	陈水扁，沈富	赖清德	绿营	约24万人	进步主义 自	06-2655120		台北市中正区北平东路	民主进步党的组织
中国国民党	中国国民党（Kuomi	1894年11月24日	孙中山、黄兴	朱立伦	蓝营	约37万人	时事、党政	(02)87711234		台湾省台北市中山区八	中国国民党的党务
台湾民众党	台湾										
中国民众党	中国										
无党团结联盟	无党										
工党（台湾）	198										
中华统一促进党	为促										
台湾基进	台湾										
台湾团结联盟	台湾										
社会民主党	社会										
亲民党	亲民										
台湾劳动党	劳动										

关注领域	电话	邮箱	地址	组织架构	职能/主张	对华态度	社交账号	组织特点(字符串)	坐标(小数	别名(字符	英文名(字符串)
政治、民主	(02)2756-2222		台北市松山区光复	郁慕明曾经把新	新党主张清廉制	亲华	00044581712	中华民族主义、中国	121.52706		New Party (Taiwan)
进步主义 自	06-2655120		台北市中正区北平	民主进步党的组	如果国共片面和	反华	00044392990	“台独”、“住民自	121.52007	民进党	Democratic Progressi
时事、党政	(02)87711234		台湾省台北市中山	中国国民党的党	依据《中华民国宪法》，		00044365992		121.54176	国民党	Kuomintang of China
民主、自由、多元、开放、法治、人权			台北市中正区杭州	台湾民众党自命	民众党宗旨为，以台湾的		00044175652		121.52007		Taiwan People's Party
			基隆市信义区忠孝里深溪路36巷70弄23号					信仰国家社会主义，	121.77229		
			台北市大同区南京西路344巷25号6楼					自由保守主义	121.51392	无盟无党联	Non-Partisan Solidar
			台南市新化区民治路43巷21号					两岸统一、社会主义	120.34361		
			新北市中和区建康路21号7楼					和平统一、一国两制	121.49308		Unionist Party(Chine
										基进党TSP	Taiwan Statebuilding
										台联党	Taiwan Solidarity Un
										社民党	Social Democratic Pa
											People First Party
											The Labor Party of T
											The People Union Par
											Green Party Taiwan
										时力（NPP	New Power Party
										维新党	Taiwan Renewal Par

姓名	婚姻状况	简介	出生日期	性别	党派	政治倾向	受教育程度	年龄	民族	职务	工作单位	教育经历
邱国正	已婚	邱国正（1	1953年4月12日	男	无党籍			70		台湾国防部部长	台湾国防部	台北市立莹
锺树明		锺树明（1	1964年	男				59		第10任陆军司令	台湾陆军司令部	
陈建义		陈建义196	1968年	男						陆军司令部参谋长	台湾陆军司令部	
敖以智		敖以智，中华民国海军中将，		男						海军第1副司令	台湾海军司令部	
刘任远		刘任远（1	1963年	男			硕士毕业	60		空军司令	台湾空军司令部	
孙连胜		孙连胜（1	1964年	男				59	鲁凯族	空军副司令	台湾空军司令部	
马英汉		马英汉，中华民国国军陆军中将，为第一任资通电军指挥官，					硕士毕业			指挥官	台湾资电作战指挥部	
梅家树		梅家树（1	1963年	男				60		海军司令部参谋长	台湾海军司令部	
黄佑民		黄佑民，中华民国海军中将，		男			硕士毕业			常务次长	台湾国防部	
陈忠文		陈忠文（1	1964年	男				59		陆军副司令	台湾陆军司令部	
蔡沐霖		蔡沐霖（1	1983年4月18日	男	民主进步党	绿营		40		民进党前青年发展部主任	民进党青年发展部	

涉T数据积累

类型	子类	数据量
新闻资讯及社媒数据	内容数据，关系数据	6221577+
全球涉T信源	——	10000+
重点目标社交账号	——	5000，可实现虚实映射
T知识图谱数据	——	10w以上，包括：人物共14881，三元组共182238；其他实体：56413，三元组共403242，总的三元组数：585480。
人物库	T政要画像	170
	T民众户籍数据（闭源数据）	2300w
	宗教界、学术界、商界人物等	13847
组织库	T政党资料	75个政党
	T智库名单	——
	T企业资料	1478家企业，通联方式、地址（台在大陆的企业、500强、上市企业等）
	T宗教资料	13546个宗教团体
	T民调机构资料	——
	T族群结构资料	——
	T人民团体相关资料	23785个人民团体

涉T数据积累

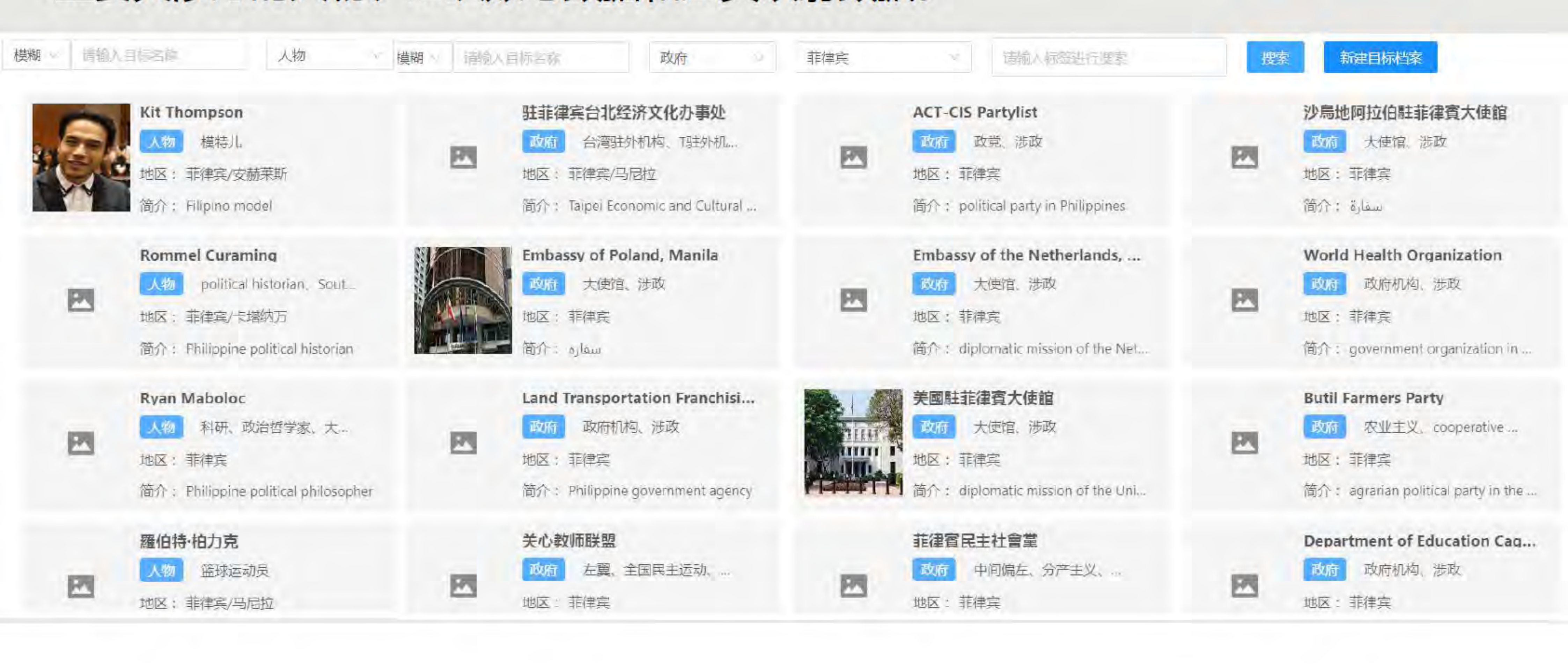
基础资料	TW选举相关资料	——
	TW政治献金相关资料	台湾大选期间台内部的政府献金，献金对象、时间、金额、来源
	美对T军售资料	——
	T在美游说相关资料	——
	美军舰自由航行资料	——
	TW稳定性评估指标数据	2006-2021年，51个指标，经济社会指标
	T社会域基础资料（人口、就业率、宗教/族群分布……）	——
	2017-2012路透社数字新闻研究所报告（包含Tw社群平台受欢迎程度排名，新闻信任度，报纸、电视、收集新闻的受众等）	
	TW民众TW人/中国人认同趋势分布	来源：国立政治大学选举研究中心
	TW民众统独立场趋势分布	来源：国立政治大学选举研究中心
	TW民众政党偏好趋势分布	来源：国立政治大学选举研究中心
	TW历年统计年鉴	——



涉南海数据

涉南海数据：人物、组织账号数据

美国、日本、越南、菲律宾、马来西亚、文莱、印尼、中国台湾等周边声索国及主要关涉方的人物、组织账号数据和虚实映射数据。



涉南海数据：政情百科数据

美国、日本、越南、菲律宾、马来西亚、文莱、印尼、中国台湾等周边声索国及主要关涉方的政情百科数据，涉及基本国情、人口、城市、宗教、军事实力、犯罪率等。

国家地区	时间	军事地位（排名	现代化水平	军事预算(单位：美	军队规模	核武器（单位
印度尼西亚	2000年		国家军由正规部	1.13E+09	484000	无核
印度尼西亚	2001年			9.19E+08	499000	无核
印度尼西亚	2002年			1.37E+09	499000	无核
印度尼西亚	2003年			2.13E+09	492000	无核
印度尼西亚	2004年			2.43E+09	492000	无核
印度尼西亚	2005年	24		2.15E+09	492000	无核
印度尼西亚	2006年	24		2.61E+09	497000	无核
印度尼西亚	2007年	13		3.35E+09	582000	无核
印度尼西亚	2008年	13		3.23E+09	582000	无核
印度尼西亚	2009年	13		3.30E+09	582000	无核
印度尼西亚	2010年	14		4.66E+09	582000	无核
印度尼西亚	2011年	18		5.84E+09	582000	无核
印度尼西亚	2012年	15		6.53E+09	582000	无核
印度尼西亚	2013年			8.38E+09	582000	无核
印度尼西亚	2014年	19		6.93E+09	676500	无核
印度尼西亚	2015年	12		7.60E+09	676500	无核
印度尼西亚	2016年	14		7.40E+09	676500	无核
印度尼西亚	2017年	14		8.80E+09	676500	无核
印度尼西亚	2018年	15		7.49E+09	675500	无核



涉Z数据

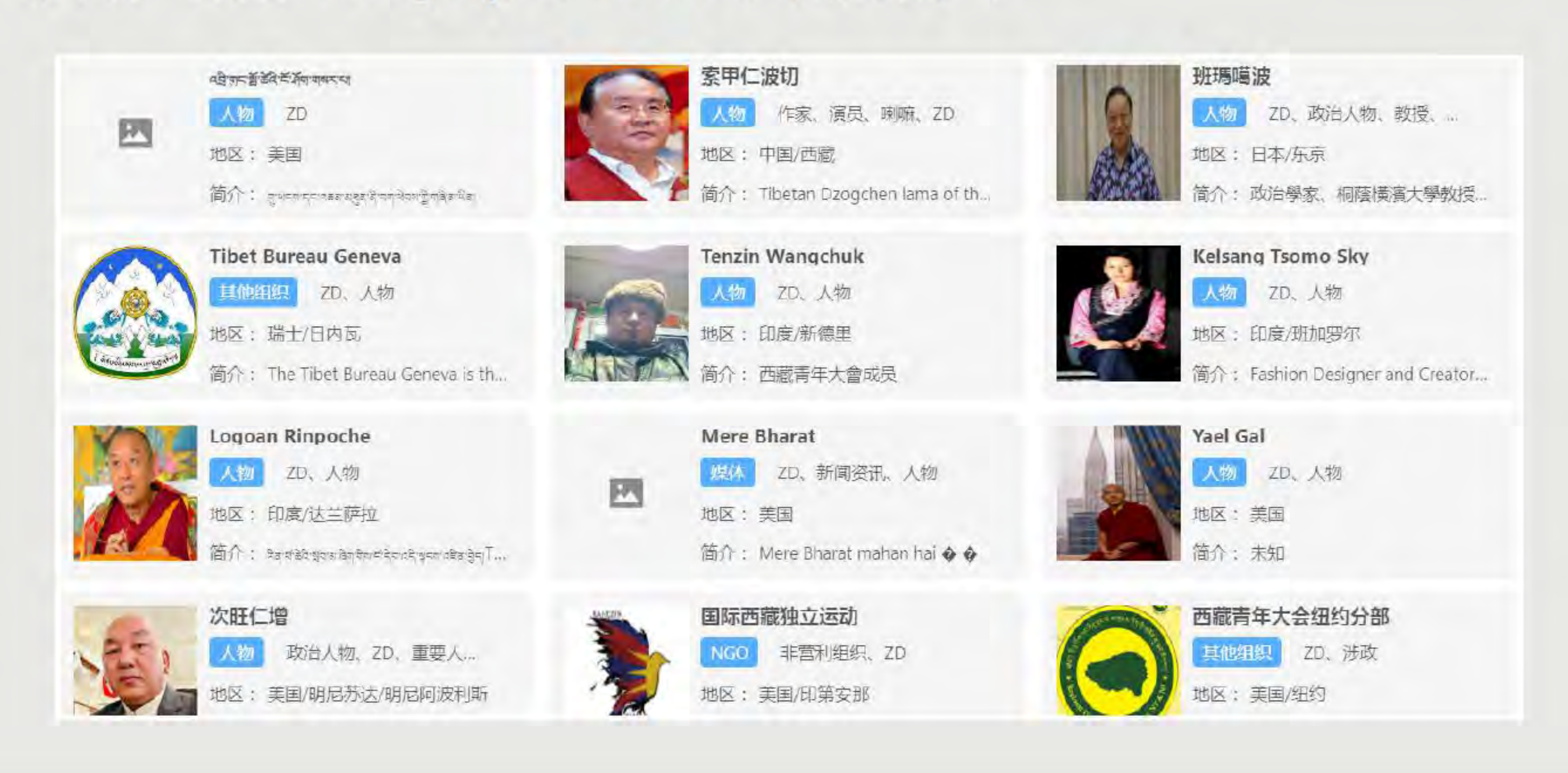
涉Z数据积累：人物组织数据

可虚实映射的涉藏人物和组织100+，含姓名、简介、性别、国籍、所在地区、出生日期、语言、外文姓名、种族民族、所属政党、出生地、职业、担任职务、就读院校、社交账号（脸书、推特、电报、telegraph、reddit、油管等）。

序号	平台	中文名	账号	链接	备注
1	推特	自由亚洲[illegible]台	RebelRFAChinese	https:[illegible]FAChinese	
2	推特	法广	RFI 華語 - 法國國際廣播電台	https://tw[illegible]com/RFI TradCn	
3	脸书	西[illegible]新闻	Tibet News	https://www.[illegible]k.com/Tibatnews/	
4	脸书	西[illegible]视台	TIBET TV		
5	脸书	自由亚洲电台[illegible]版			
6	脸书	[illegible]国之音[illegible]版			
7	推特	[illegible]尔时报中文（[illegible]尔之声）	[illegible]中文（[illegible]	https://twitter.[illegible]	
8	推特	RFI 华语 - [illegible]国国际广播电台	RFI 华语 - 法国国际广播[illegible]	https://twitte[illegible]	
9	推特	[illegible]之音争鸣论坛	美国之音争鸣论坛	https://twitter.co[illegible]china	
10	推特	DW 中文- [illegible]国之声	DW 中文- 德国之声	https://twitter.com/[illegible]	
11	推特	BBC News 中文	BBC News 中文	https://twitter.com/b[illegible]	
12	推特	[illegible]音中文网	[illegible]国之音中文网		
13	推特	[illegible]中文笔会 ICPC	[illegible]中文笔会 ICPC	https:[illegible]	[illegible]文笔会

涉Z数据积累：ZD人物及组组虚实映射数据

可虚实映射的ZD人物及组织近百个。含社工数据、人脸数据、社交账号（脸书、推特、电报、telegraph、reddit、油管等）。





6 涉J数据

涉J数据积累：JD人物组织虚实映射数据

可虚实映射的JD人物及组织近百个。含社工数据、人脸数据、社交账号（脸书、推特、电报、telegraph、reddit、油管等）。

库扎特·阿尔泰

Former Chinese Citizen | American Businessman | Harvard Business Alum | Independent Uyghur State | Founder & CEO https://t.co/ZtXr9GBeH4 | China, Uyghur, Tech | 真实姓名库扎提·买买提，原"世维会"主席热比娅·卡德尔的侄子，目前在美国活动，常与美国议员有密切接触，并时常参加欧洲议会、美国国会听证会等西方政府官方正式场合，与美国政府工作人员或其他活跃"东突"分子及"东突"组织领导人共同参加示威活动，时常举行维吾尔议题相关会议并发表讲话。籍贯：新疆乌鲁木齐，国籍：美国，宗教信仰：伊斯兰教，已婚，妻子拉希拉·阿尔泰系；教育背景：1997年至2003年，就读于乌鲁木齐市第十四中学；2018年至2019年，参加哈佛商学院高管培训课程。精通中文、英语、维吾尔语和土耳其语；职业：企业家、演讲家，办公地址 Cybertek：美国弗吉尼亚州泰森斯角琼斯布兰奇路7920号3300号办公楼（7925 Jones Branch Dr, #3300, Tysons, VA 22102, USA）；职务："美国维吾尔协会"主席、"维吾尔企业家网络"创始人、信息技术教育公司Cybertek创始人兼CEO；家庭住址：美国弗吉尼亚州麦克莱恩市风山路1327号（1327 Windy Hill Rd, Mc Lean, VA 22102）工作履历：2008年8月至2012年8月，在华盛顿大都会区的Consultant公司担任软件质量检查工程师；2012年9月至2013年10月，在美国技术公司Inovalon任自动化测试工程师；2013年10月至2014年3月，在总部位于加拿大蒙特利尔的跨国公司CGI集团任自动化测试工程师；2014年3月至2014年10月，在跨国管理咨询公司埃森哲（Accenture）任自动化开发员；2014年10月至2016年2月，在美国房屋抵押贷款公司房利美（Fannie Mae）任自动化开发员；2015年8月至今，创办信息技术教育公司Cybertek并担任首席执行官；2019年10月29日，当选"美国维吾尔协会"主席。

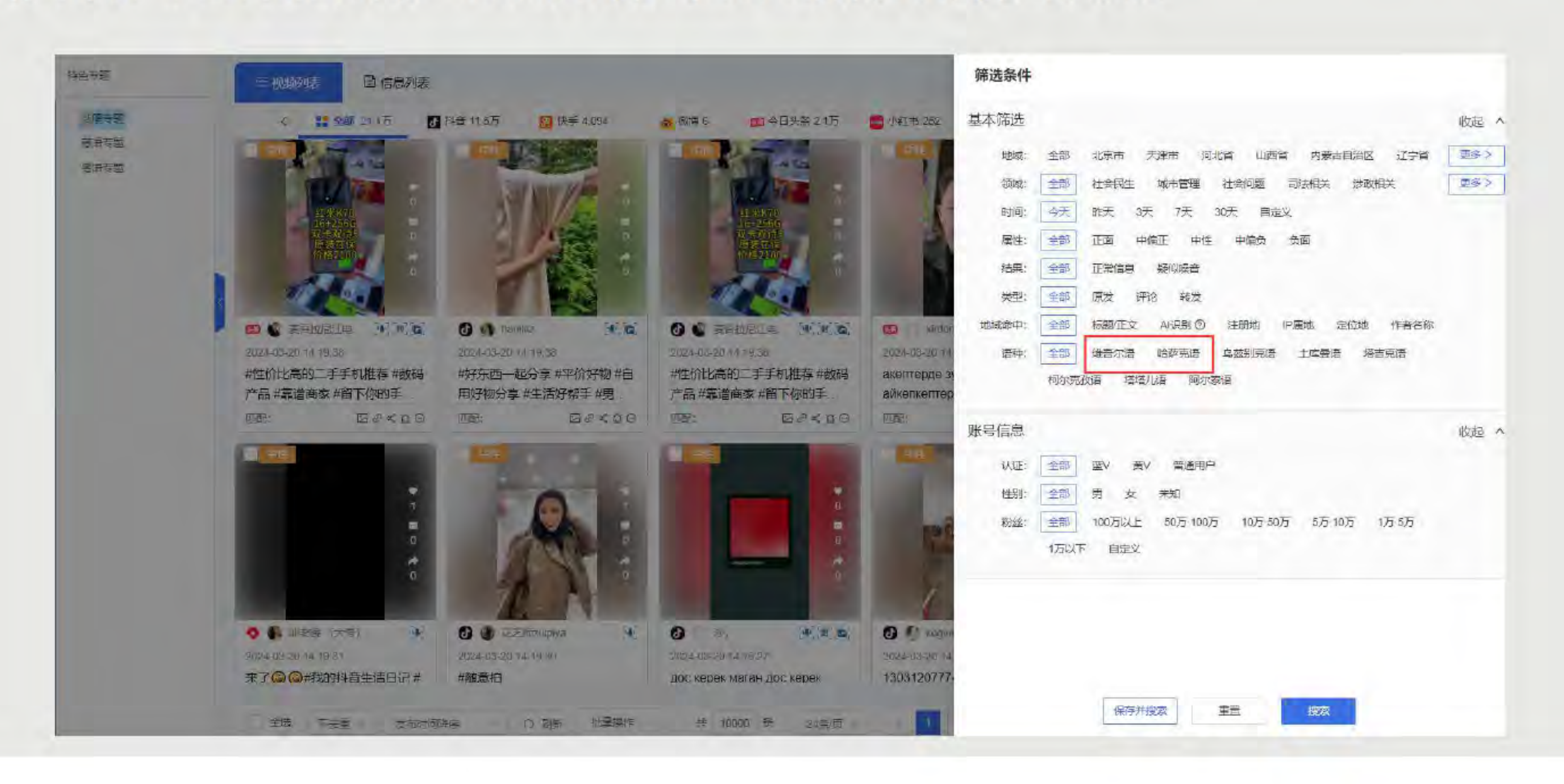
目标类型 人物　　目标标签 涉疆 疆独

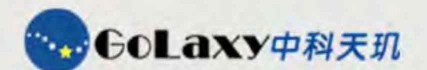
基本信息

性别	男	国籍	美国	所在地区	美国
出生日期	未知	语言	未知	外文姓名	Kuzzat Altay
种族民族	未知	所属政党	未知	入库时间	2023-08-23 15:48:10

涉J数据积累：8类小语种语音识别的舆情数据

涉疆信息全量语音识别，支持维吾尔语、哈萨克语、乌兹别克语、土库曼语、塔吉克语、柯尔克孜语、塔塔儿语、阿尔泰语8类语音识别。







THANK YOU

电话：400-901-9755

官网：http://www.golaxy.cn/

地址：北京市海淀区西三旗中科智汇工场N1楼301号



面向争夺国际话语权和维护国家意志需要，中科天玑团队投入近 50 余人，基于在舆情、情报分析领域多年积累的学术与技术成果，研发出可服务国家层面的认知域对抗、引导的智宣 GoPro 系统。

GoPro 系统部分技术成果已在国务院港澳办、安全部、联参情报局、战支一线部队等部门部署应用，并为应用部门在回应党中央重大关切，诸如香港问题，台海问题，中美围绕疫情、贸易、高科技、大选等展开的博弈，印度地缘政治冲突等方面，提供了态势感知、事件分析、舆论引导等技术支持。

下面给各位领导详细介绍 GoPro 系统。

首先，GoPro 系统综合不同应用单位的工作需要，基于天玑 WDE 平台，从传统媒体、社交网站、即时通信等 800 万个信源，广泛采集香港、台湾、美国、印度等地区的政治、军事、外交、经济数据，并每日实时聚合出不同区域的热门和敏感政治事件，深度分析与该事件相关的话题、媒体和人物，挖掘媒体和相关人物所持立场与观点。

其次，GoPro 系统基于积累十年的情报分析技术和 GoIN 系统，对政治事件幕后组织、机构、人物进行精准画像，并通过广谱关联技术实时锁定目标，以近期香港国安法和蓬佩奥拟访台事件为例，从全网监测的 966 万 twitter 用户和 18 万香港用户中，锁定与这两起事件相关的 3000 多

名主要参与人物、100 多名网络意见领袖和 20 多家新闻媒体。锁定目标后，快速回击目标对象的黄论、谬论，及时正本清源。

第三，在事件的舆论征候和特定区域的风俗习惯等基础信息之上，一是通过研判对方涉我言论的真伪和逻辑漏洞；二是寻找对方内部分化和极化观点，比如用马斯克对中国抗疫的言论反驳美西方国家对我的抹黑和污蔑，实现以己之矛攻己之盾；三是利用人工智能技术如自动联想、素材借鉴、文案衍生、长文本生成等方式辅助创作中国抗疫好故事。

最后，非受控环境下的智能推送，综合解决一体化指挥调度、智能化传播导入、科学化评估反馈、拟人化生存等技术难题。视频中向各位领导展示，利用智能推送功能，可实现跨 Twitter、Facebook 等平台同时对目标对象进行实时、高可达、浸入式的舆论引导。

天玑团队自研的 GoPro 系统，目前已具备政治态势感知、实时锁定、高质量内容创作、快速回击等能力，已在国家相关部门产生一定政治效应。未来我们期盼在科学院领导和所领导的大力支持下，进一步加强与中央网信办、中宣部、军科委、安全部等部门的合作，将 GoPro 打造成为能真正自动讲好中国故事、发出中国声音、扩大中国影响力的技术平台，为早日实现“东风压倒西风”提供全方位的技术支持。

会议纪要

会议主题：纵横系统工作进展汇报沟通会

会议时间：2023 年 5 月 11 日 下午 2:00

会议地点：通研院林萃桥办公区一层大会议室

参加人员：郭杰，三家课题相关人员

苏鹏， 课题五，南京信息技术研究院

一、课题一（王清林，刘杨昊，黄振明，王洪航，隋雨恒）：

1. 汇报进度：完成全部内部编码，完成联调（90%）。
2. 下两周迭代：调试数据，集成测试，沟通外部接口。
3. 播放系统演示视频，大家评审。
4. 会议输出：

（一） 郭所反馈：郭所了解了课题一整体工程进度（在计划内）。同时要求课题一刻盘当天的演示视频给他，已完成。

（二） 确定下周对外沟通会议议程：

1. 接口课题三：下周一下午下午（5 月 15 日），去清华，讨论仿真推演数据需求和方案。
2. 接口课题二：下周四下午（5 月 18 日），地点另行协商讨论对接接口（1.敏感事件，2.隐含关系，3.为舆图提供活数据）

（三） 苏老师反馈：

Q：事件生成的关键词逻辑可以支持多关键词间的“与、或、非以及分组逻辑”。

A：回去调研相关算法部分。

（四） 课题三反馈：

Q：数据资源监控功能没有体现。

A：数据资源监控功能做在运维视角子系统里面，不在当前面向业务人员视角的子系统里面。天玑有相关设计，未开始。

（五） 视频演示反馈或者发现的问题：

1）事件：相关推文的“暂无信息”

2）群组：按什么维度去划分,存在几个建群组的属性，

3）人物画像：存在字符集问题；有多少个维度

4）平台和语种：语种问题不大，单个平台到多个平台画像对齐没有设计

5）降噪处理：事件关键词组合过滤文本

6）新闻数据 时间问题（2023 年 8 月 1 日的新闻）

二、课题二（仝春艳，闫洲）：

1. 系统开发进展：

特定 HL（黑料）曝光：基于内容的隐含线索挖掘

涉 H 议题引导：摇摆人（精英人物、普通网民）

摇摆人受众精准定位：精英人物，基于利益立场导向

普通民众，基于情感态度导向

XJ（选举）舆论引导：最复杂的，两党中的精英之争

XJ 舆论场引导：两党汇总的精英之争->选举舆论场态势

技术。

感知->下发合理的任务目标->基于 XJ 影响的策略制定

评估 XJ 舆论场两党各自的主要影响因素:确定主要人物对华态度->评估精英动员的影响能力->确定特定议题对候选人的影响路径

任务评估：多次投放，多次评估，目前整体做评估

用户 zz 心理画像模型：

zz 立场模型

大五人格模型

舆图-议题-党派分布：效果图 还未接入真实数据

舆图-议题-社会阶层分布

账号可以定位地区上

人物视图也加进来了

议题下面有评论的数据 ：前两百条左右

地图数据：采集不到，使用框架进行集成，开源的地图框架 openMap?

年龄分布：用模型去推断年龄

立场：在一个议题下，以账号视角去看，三种立场：支持，反对，中立。

语种：目前能够支持中英文，模型好说，训练集比较难找

图数据库：neo4j

2. 会议反馈

郭所：关注现有设计前提下实际采集来的数据支持能力如何，

系统要能数据落地，课题二反馈目前有推断数据。

苏老师：可以基于原始数据进行属性补全

对课题三数据库设计反馈（训瀚）：重构数据库部署设计，三家数据库分隔开。各个课题分库：读写分离。

对课题一数据采集有特定要求（接下来会提出）：

议题，带评论的数据，前 200 条即可

地图，采集有困难，目前用的开源数据）

三、课题三（高云鹏，徐文博，刘训瀚等四人）：

1.完成素材发现的设计和功能实施

2.机器人自动化注册脚本以及接口封装（短期：半自动人工验证）

3.机器人账号的作战任务的设计，日志视图

4.新增 200+ Twitter 草根账号的数据

5.仿真推演模型的实施方案

账号养殖情况：

1. twitter：婴儿期大 v：10 婴儿期普通：50 草根：600

2. reddit：20 个草根

3. facebook：完成基础调研工作：不好找突破口

4. 购买境外服务器，用 IP 池，精养账号绑定固定 IP，炮灰账号 IP 在州范围内流动。

5. 目前用了两天服务器，采用 Selenium Webdriver 模拟

技术。

账号养殖问题：

1. 账号养殖：模拟器 网页端 培育：手机版、模拟器
2. 使用邮箱账号，后面有可能要求绑定手机号
3. 自动化脚本：当前围绕网页，调研模拟器
4. 没有动作，被锁的概率：5%
5. 邮箱：目前用的开源邮箱，自己搞一个域

下两周任务：

1. 仿真推演

作战投放

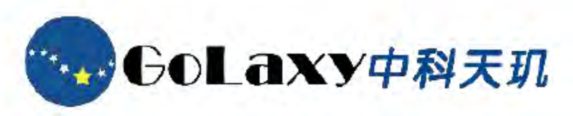
Facebook 平台突破方案,交付多平台账号养殖

权限方案设计

郭所反馈：素材发现不仅满足于采集形成的素材，尽量支持语义搜索形成轻量级的素材挖掘渠道。

高老师回复：已实现采集回来的素材（新闻），工程易实现支持语义搜索的素材挖掘，在模型开发过程中。

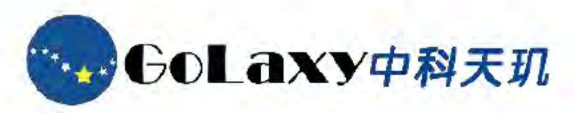




中科天玑数据科技股份有限公司

技术产品体系





公司战略和产品定位

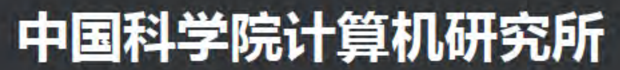
战略定位





1956年成立

- 从事计算机科学技术综合性研究的学术机构
- 中国计算机事业的“摇篮”
- 国家高性能计算机研发基地



- 中国计算机事业的“摇篮”
- 孕育创建联想、曙光、天玑等20多个各领域高新技术企业
- 1958年 中国第一台数字电子计算机 -10
- 22位院士工作或学习过
- 孕育创建联想、曙光、天玑等20多个各领域高新技术企业

联想

曙光

龙芯

天玑





2010年成立

- 中科院计算所孵化的高新技术企业
- 以大数据智能计算引擎、人工智能算法为核心
- 服务于党政军和教育、金融等行业

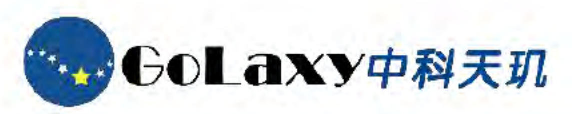








战略定位



大数据分析系统国家工程实验室

- 突破大数据分析技术瓶颈，研制第三代大数据分析软件栈
- 2017年发改委批复依托计算所成立；李国杰院士任主任，程学旗副所长任常务副主任
- 共建单位：国科大、网络中心、中科天玑、曙光、国创科视

中科院网络数据科学与技术重点实验室

- 中科天玑团队的重要组成部分
- 先后承担了多项国家重点863、973、科技攻关等重大课题项目，科研成果多次获得国家及省部级奖励

中科天玑

计算所大数据方向唯一产业化平台

计算所&地方分所

- 烟台网络技术研究所：80人
- 郑州大数据研究院：50人
- 福州数据应用技术研究院：40人
- 厦门。。。

中科天玑研发中心

- 75%占比的技术研发团队
- 提供党军政、教育、金融等领域的大数据整体解决方案的应用



大数据分析技术领域“国家队”，硕士、博士100+人，科研人员400+人

战略定位



服务于国家安全、国家战略

三大目标市场：党、政、军

二大技术应用：舆情、情报大数据分析

一个核心技术系统：天玑数据大脑

情报大数据分析
第一品牌

天玑技术产品体系及产品家族



网信行业整体解决方案
舆情监测系统3.0、XWTK系统2.0、指挥系统
重点人员管理系统、属地管理系统

安全行业整体解决方案
大魔镜系统3.0

JG行业解决方案
开源情报监测分析系统GOIN

舆情监测分析系统 | **智能情报分析系统**

金融舆情系统解决方案
金融监管大数据分析系统产品

教育行业解决方案
大数据教学实验平台

通用组件产品

自然语言处理 NLU/NLP	开放知识系统OKS	图计算引擎 SQLGraph	内存数据库 GOLAXDB	。。。

基础平台解决方案
超融合架构
大数据一体机

基础平台产品（天玑数据中台）

数据采集服务平台（WDE）	大数据存储管理平台（BDE/C）	可视化智能机器学习系统（BDA）

自主可控大数据基础平台
BDA600-L龙芯
BDA600-F飞腾

天玑数据大脑技术生态系统



知识图谱

领域知识

公共知识

网信应用 | JG应用 | 安全应用 | 教育应用

天玑数据大脑

算法分析服务

数据治理加工

数据存储管理

数据标准

行业数据模型

数据标准

互联网开源数据采集

工业物端数据接入

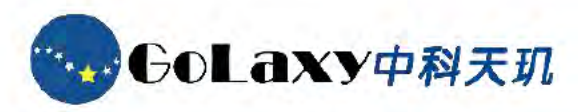
Internet

IoT

天玑数据大脑提供从数据到价值的数据服务生产全链条

1. 数据模型能力
2. AI算法模型能力
3. 行业应用能力

中科天玑中台体系



实际交付系统




根据项目差异化需求完善

业务中台

业务系统产品基线和通用业务模块




根据数据服务API构建完整产品

数据中台

采集的基础数据加工处理后提供服务



根据已有的算法和框架对数据进行加工分析

算法中台

面向各类数据和场景的算法能力以及服务封装

技术中台

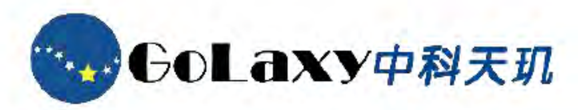
通用的大数据存储管理、框架、引擎





基础平台产品

基础平台产品线



天玑数据大脑，为各行业产品和应用提供数据中台支持

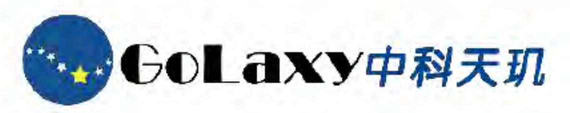
- 可视化智能机器学习系统（BDA）
- 大数据存储管理平台（BDE）
- 数据采集服务平台（WDE）

衍生产品形态

- 自主可控大数据一体机
- 超融合架构大数据一体机



数据采集服务平台（WDE)



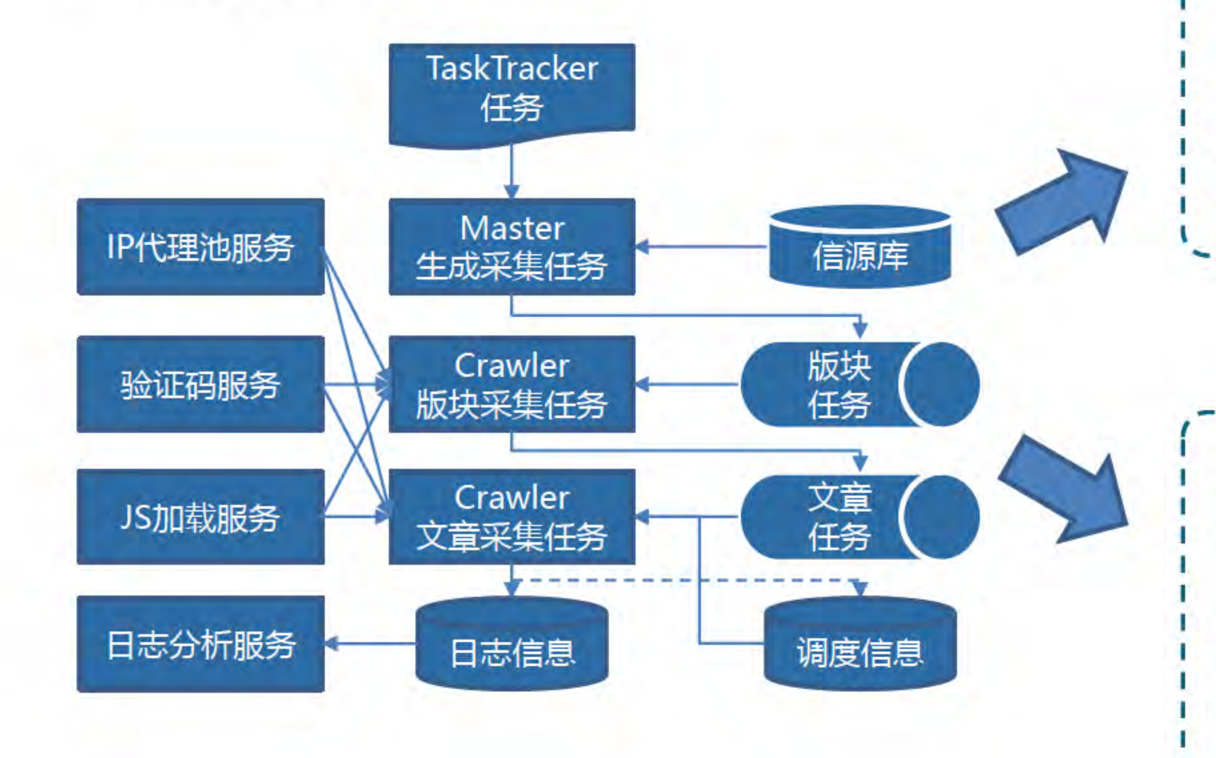
- 面向政府监管、安全、军工、金融、教育、企业等领域提供数据服务
  - 定制化采集与分析，包含实时数据采集、历史数据回溯、热点数据、数据分析报告，实现数据的按需获取
  - 以多种API接口或文件的方式进行推送
  - 平台覆盖境内和境外海量数据，支持国内主流新闻媒体、微博、微信、APP客户端，以及论坛、博客、贴吧、问答社区、短视频等通道，以及境外媒体、facebook、twitter等
  - 全天采集数据量在8000万-1亿



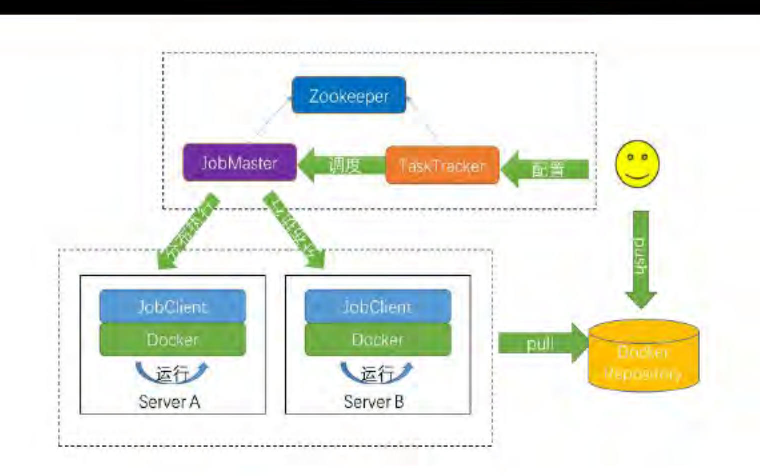
数据采集服务平台（WDE）



- 产品架构



分布式调度框架

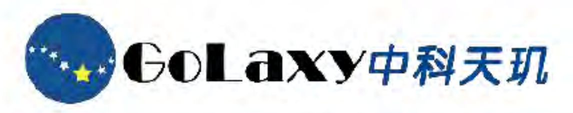


采集服务框架





大数据存储管理平台（BDEC）



- 大数据存储管理平台（BDE）是企业级大数据一体化应用开发基础技术平台，将容器云平台与大数据系统相结合，实现了云+大数据+智能服务的产品技术融合，为用户提供构建大数据业务系统所需的完整的基础平台能力
- 基于容器云环境，提供了IaaS+PaaS+SaaS多层服务能力，不同层级的组件和服务统一运行于容器云支撑平台，相互共享资源，资源可以依赖运行情况弹性伸缩，提高整体运行周期的资源利用率
- 覆盖数据存储、分析挖掘和应用构建三方面需求，是一个具备大数据系统全栈研发能力的一站式、标准化、可视化的智能大数据统一研发平台

大数据存储管理平台（BDEC）



应用服务SaaS
政府行业
军工行业
金融行业
教育行业
安全行业

数据平台
数据资源体系
元数据管理
资源目录
数据浏览分析
数据汇聚/抽取/转换/融合
数据资源池
数据流转
统一SQL
大数据组件
检索
文档
键值
计算框架
......

分析服务组件
可视化分析工具
组件服务管理
实体识别
关键词
实时聚类
视频分类
图片检索
情感分析

应用构建平台
Web中间件
虚拟路由管理
服务流量管理
统一认证
统一权限管理

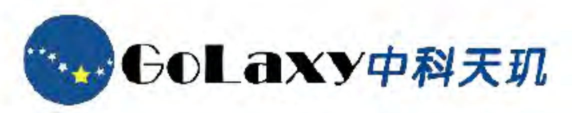
应用管理
应用市场
自动发布
自动部署
自动构建
DevOps流水线

支撑管理
微服务管理
灰度发布
运维管理
智能告警
可视化监控
日志采集
日志管理
安全管理
权限管理

容器云管理平台
容器管理
镜像服务
应用管理
存储服务
网络管理
统一配置
服务管理
资源管理
Docker+Kubernetes

混合 IaaS
私有云
物理机
VmVare
Openstack
公有云
阿里云
腾讯云
Amazon

可视化智能机器学习系统（BDA）



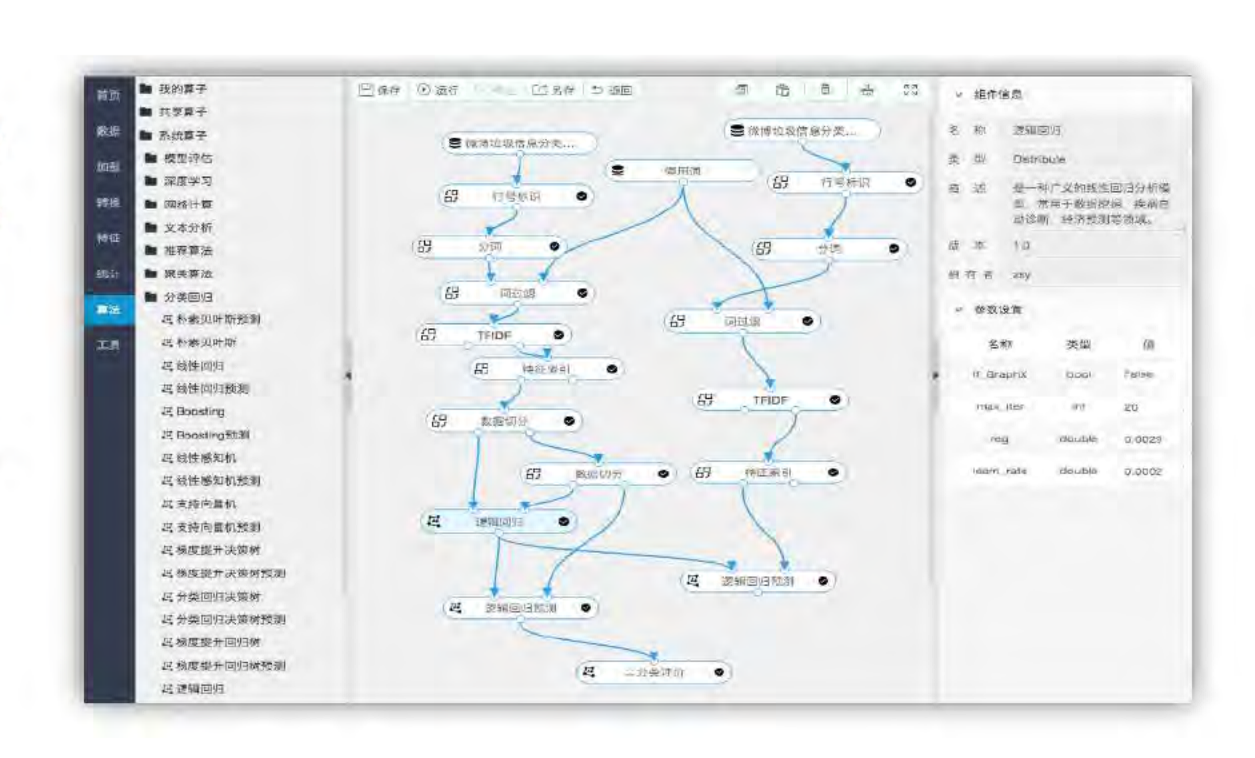
- 可视化智能机器学习平台是集数据预处理、模型训练、预测、评估为一体的**一站式大数据智能分析挖掘平台**，可帮助用户**快速构建**和**持续优化**大数据分析模型，挖掘大数据资产价值

✓ 基于全流程的可视化大数据分析平台，无需编程，快速实现大数据处理流程，降低使用门槛，提高工作效率

✓ 算法训练平台，迭代优化

✓ 针对不同业务场景复用和针对性优化算法模型

✓ 训练好的算法提交发布，形成服务



可视化智能机器学习系统（BDA）



系统管理
- 用户权限：用户管理、权限管理、角色管理
- 日志管理：模型训练日志、系统操作日志
- 监控管理：集群监控、应用监控

计算资源：CPU、GPU

模型应用：模型共享、模型发布、在线应用、离线分析、行业模型、......

模型构建
- 自定义：模型上传、参数定义、在线编程
- 框架集成：机器学习、深度学习、......

模型训练：模型提交、模型优化、日志分析、模型保存、参数调整、结果预览、状态监控、模型运行、......

模型评价：二分类评估、多分类评估、困惑度评估、均方根误差、R平方估计、......

统一模型调度接口

模型计算引擎：任务执行引擎、机器学习引擎、深度学习引擎、图计算引擎、资源调度接口

数据管理：数据接入、数据处理、数据清洗、数据采样、格式转换、数据存储

数据源：外部数据源、互联网、舆情、金融、安全

可视化：流程可视化、数据可视化、模型可视化、Dashboard

大数据分析算法库（BDA Lib）

图形化交互界面（BDA Studio）

分布式计算框架（BDA Infrastructure）



自主可控大数据一体机



自主可控大数据一体机（Golaxy Smart and Security Data Appliance，简称SDA）是中科天玑公司面向党政军客户推出的系列产品，针对客户对大数据和自主可控方面的迫切需求，提供数据离线分析、实时分析、数据仓库、机器学习、数据挖掘、数据安全、商业智能等通用基础平台支撑，兼容支持龙芯+中标麒麟、飞腾+银河麒麟以及国产数据库、中间件等自主可控产品，具有高性能、高可靠、高安全、高扩展、易用性、高性价比等特点和优势，为客户提供自主可控、安全可靠、智慧敏捷的一体化大数据解决方案。

序号	产品型号	典型配置
1.	SDA60000L（龙芯）	服务器：处理器龙芯3B3000及以上，内存128GB及以上。
		最少6台，支持线性扩展。
		操作系统：中标麒麟军用服务器操作系统v5.0，内核3.10
		数据库：国产数据库
		大数据平台：中科天玑大数据平台BDE v2.0
		适用场景：应用中间件、数据库、负载均衡、安全管理、运维管理、安全可信信息系统等
2.	SDA60000F（飞腾）	服务器：处理器飞腾FT2000及以上，内存128GB及以上。
		最少6台，支持线性扩展。
		银河麒麟服务器操作系统v4.0，内核4.4.131
		数据库：国产数据库
		大数据平台：中科天玑大数据平台BDE v2.0
		适用场景：计算密集型和I/O密集型，业务密集型应用、高性能计算、安全可信信息系统、云计算等

大数据一体机SDA60000L





- 硬件：同品牌、同型号、同配置
- 软件：同操作系统、同JVM虚拟机、同系统账户体系

序号	分类	类型	技术规格	数量（台套）	备注
自主可控大数据一体机SDA60000L	硬件	服务器	机架式2U龙芯服务器，两颗龙芯3B3000系列处理器（4核 1.45GHz），内存64G，一块 960G SSD，六块2.5″ 1200G/10000RPM/12Gb/SAS，一块Intel 网卡:PCI-E 双万兆光口网卡 半高（含模块），一块RAID卡:12G 3108芯片 八口 R0 1 5 6 10 50 60 2GB缓存 半高，2U 3.5寸8盘位类型，板载Intel 四千兆网口，最多支持4个PCIe槽，含导轨，建兴内置DVD光驱。1块2U 550W AC 白金电源(冗电源)，一套中标麒麟操作系统V7.0（服务器版）（试用版）。服务：提供三年原厂服务，硬盘三年免回收服务。	6	长城超云国产龙芯服务器R3210 L10
		业务交换机	业务接口：24个万兆光口数（SFP+），4个100G光口数（QSFP28） 管理接口：1 个RJ-45 端口，1 个RJ-45 Console 口，1 个USB 端口，1 个Mini USB Console口 MAC：支持128K MAC 表项 时延：<2us	1	迪普自主可控万兆交换机LSW6600-24XGS4CQ-G
		管理交换机	业务接口：28 千兆电+4 千兆光（SFP）+4万兆光（SFP+）， 管理接口：1 个RJ-45 端口，1 个RJ-45 Console 口，1 个USB 端口，1 个Mini USB Console 口 扩展槽位：1个 MAC：支持64K MAC 表项，支持静态MAC、动态MAC、黑洞MAC、源MAC 地址过滤	1	迪普自主可控千兆交换机LSW5662-28GT4GP4XGS-G
		KVM		1	国产KVM
		机柜		1	
	软件	操作系统	国产操作系统	6	中标麒麟服务器操作系统v7.0
		数据库	国产数据库	1	达梦、金仓等
		中间件	国产中间件	1	中创、东方通、金蝶等
		大数据平台	国产大数据平台	6	中科天玑大数据平台BDEv2.0
		Python、Java运行开发环境		6	操作系统自带
		浏览器		6	操作系统自带

大数据一体机SDA60000F





- 硬件：同品牌、同型号、同配置
- 软件：同操作系统、同JVM虚拟机、同系统账户体系

序号	分类	类型	技术规格	数量（台套）	备注
自主可控大数据一体机SDA60000F	硬件	服务器	机架式2U飞腾服务器，一颗飞腾FT2000+（64核 2.2GHz），内存128G，一块 960G SSD，六块2.5″ 1200G/10000RPM/12Gb/SAS，一块Intel 网卡:PCI-E 双万兆光口网卡半高（含模块），一块RAID卡:12G 3108芯片 八口 R0 1 5 6 10 50 60 2GB缓存 半高，2U 3.5寸8盘位类型，板载2个千兆网口，独立IPMI管理接口，含导轨，建兴内置DVD光驱。1块2U 550W AC 白金电源(冗电源)，一套银河麒麟V4.0服务器版(试用版)。 服务：提供三年原厂服务，硬盘三年免回收服务。	6	长城超云国产飞腾服务器R2210 F11
		业务交换机	业务接口：24个万兆光口数（SFP+），4个100G光口数（QSFP28） 管理接口：1 个RJ-45 端口，1 个RJ-45 Console 口，1 个USB 端口，1 个Mini USB Console口 MAC：支持128K MAC 表项 时延：<2us	1	迪普自主可控万兆交换机LSW6600-24XGS4CQ-G
		管理交换机	业务接口：28 千兆电+4 千兆光（SFP）+4万兆光（SFP+）， 管理接口：1 个RJ-45 端口，1 个RJ-45 Console 口，1 个USB 端口，1 个Mini USB Console 口 扩展槽位：1个 MAC：支持64K MAC 表项，支持静态MAC、动态MAC、黑洞MAC、源MAC 地址过滤	1	迪普自主可控千兆交换机LSW5662-28GT4GP4XGS-G
		KVM		1	国产KVM
		机柜		1	
	软件	操作系统	国产操作系统	6	银河麒麟服务器操作系统v4.0
		容器软件	国产容器软件	1	中科天玑、银河麒麟等
		数据库	国产数据库	1	达梦、金仓等
		中间件	国产中间件	1	中创、东方通、金蝶等
		大数据平台	国产大数据平台	6	中科天玑大数据平台BDEv2.0
		Python、Java运行开发环境		6	操作系统自带
		浏览器		6	操作系统自带

产品特点



- ✓ 大规模并行计算能力
- ✓ 海量异构数据存储能力
- ✓ 大数据实时分析能力

高性能

- ✓ 通用硬件集群
- ✓ 降低软件成本
- ✓ 降低集成交付成本
- ✓ 降低运维成本

性价比

- ✓ 采用冗余数据存储方式
- ✓ 利用Hadoop已有的容错机制，保证节点故障自动恢复和数据不丢失

高可靠

易用性

- ✓ 一键式安装部署
- ✓ 一体化集成交付
- ✓ 一站式解决方案
- ✓ 全方位专业化服务

高安全

- ✓ 全自主可控平台
- ✓ 集中式安全框架
- ✓ 数据完整性保证

高扩展

- ✓ 支持扩展至1000节点以上
- ✓ 支持快速扩容到PB级存储以上





通用组件产品



通用组件产品



智能认知

- NLU/NLP
- *图像视频分析*
- *智能搜索*

知识计算

- 开放知识系统OKS

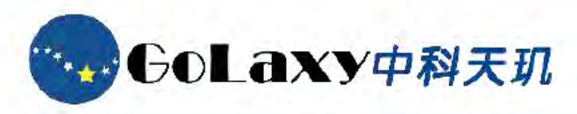
存储和算法引擎

- SQLGraph
- 内存数据库
- *机器学习引擎*

应用支撑组件

- 统一权限管理
- 可视化

自然语言理解组件（NLU）



- 自然语言处理数据处理分析组件，包括标注、抽取、链接、理解个层次十余个自然语言模型组件

✓ 覆盖自然语言理解全体系

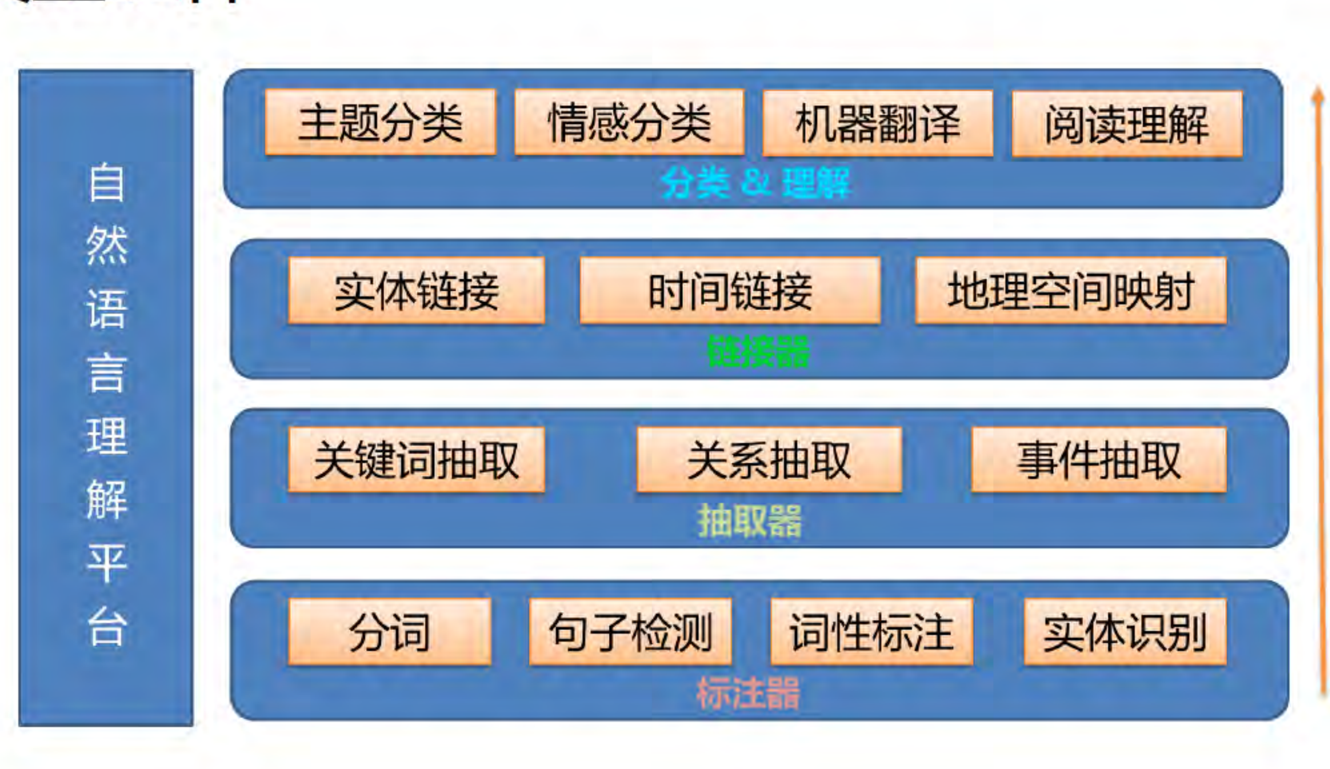
✓ 数据驱动+领域知识+机器学习

✓ 最前沿的深度学习模型、BERT-style预训练语言模型

✓ 通用/专业领域兼顾

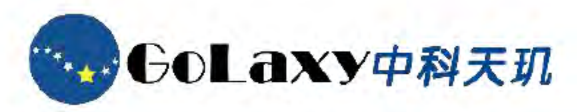
✓ 支持不同数据格式

✓ 基础库/工具级产品/在线服务





自然语言理解组件（NLU）



- 与学术界先进算法相比性能占优或相当

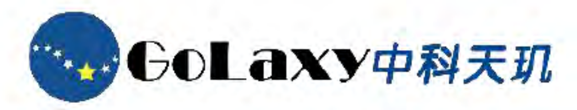
模型/数据集	词性标注	实体识别	主题分类	实体链接	事件抽取	机器翻译
	CoNLL 2003	CoNLL 2003	AG_News	ACE 2004	ACE 2005(trigger)	AI Challenge 2018
CRF (2015)	94.42	81.86	-	-	-	-
BiLSTM-CRF (2015)	**95.15**	84.18	-	-	-	-
AGDISTIS (2014)	-	-	-	66.4	-	-
NCEL (2018)	-	-	-	**88.1**	-	-
Char-CNN	-	-	0.8702	-	-	-
LEAM (2018)	-	-	0.9175	-	-	-
DMCNN (2016)	-	-	-	-	73.5	-
JMEE (2018)	-	-	-	-	75.9	-
GNMT+spm (2016)	-	-	-	-	-	**25.22**
天玑NLU	94.19	**90.59**	**0.9206**	87.3	**78.4**	24.56

- 在某军方项目POC测试中（参与方科大讯飞、拓尔思等）综合成绩第一，6项自然语言处理组件测试4项第一。

- 多年技术积累，在历次国际权威评测中名列前茅

关键技术	国际权威评测成绩
用户画像	SMP CUP2017评测排名第一
知识融合	TAC2016国际评测排名第一
知识发现	TAC2016国际评测单项排名第一
中文繁体纠错	SIGHAN2015国际评测排名第一
知识融合	TAC2014国际评测排名第一
知识发现	TAC2014国际评测单项排名第一
自动文摘	TREC2013国际评测排名并列第一
棒性检索技术	TREC2013国际评测排名并列第一
关联实体检索技术	TREC2009国际评测排名并列第一
动态文摘技术	TAC2008国际评测排名第一
企业专家检索技术	TREC2007国际评测排名第一
新信息发现技术	TREC2004国际评测排名第一
中文词法分析技术	SIGHAN2003国际评测排名第一

开放知识系统（OKS）



- 丰富的知识抽取算子库，灵活可定制
- 国际领先的关系推断算法
- 友好的知识标注工具
- 特色的行业知识图谱案例
- 支撑领域知识图谱的快速构建

实体抽取、实体链接、关系抽取等多项关键技术**国际排名第一**

知识服务与应用
知识问答 | 个性化推荐 | 语义搜索 | 知识浏览 | ......

知识计算
属性计算 | 关系计算 | 实体计算 | 概念计算 | 事件计算 | .....

知识可视化
知识评估

知识融合
事件链接 | 事件去重
关系去重 | 关系对齐 | 实体匹配
实体消歧 | 实体链接 | 概念映射

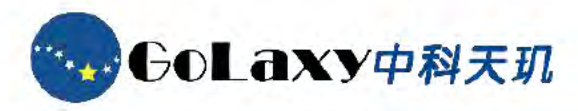
知识标注
规则定义
字典匹配
文档标注

知识抽取
概念 | 实体 | 属性 | 关系 | 事件
算法模块 | 配置模块

数据输入
类型系统 | 输入文档

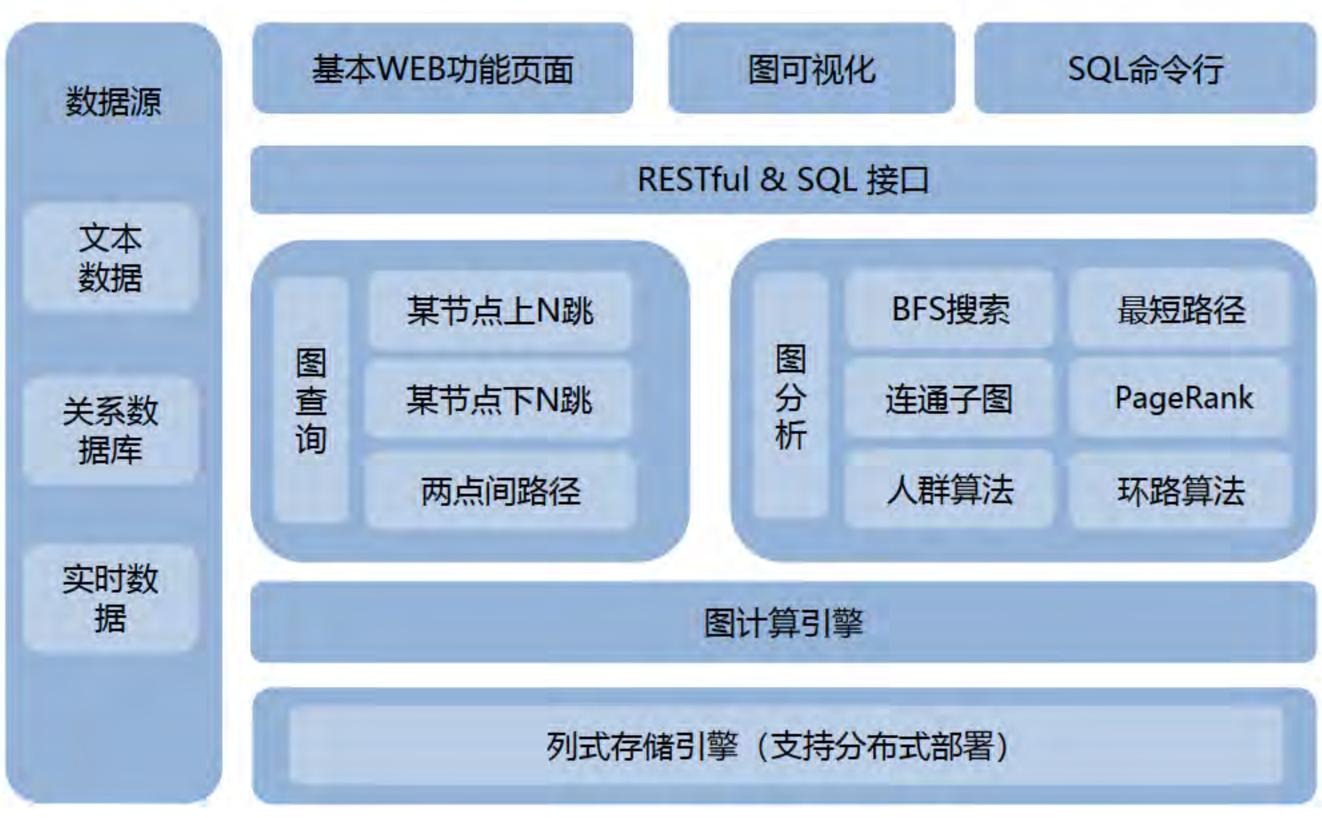
知识库 (OpenKN)
事件
关系
属性
实体
概念

图计算引擎（SQLGraph）



- SQLGraph图计算引擎是一项快速、高效的图分析数据库服务，可快速构建和高效分析高度关联的应用数据
- 可广泛应用于社交关系分析、推荐、精准营销、舆情及社会化聆听、信息传播、防欺诈等具有丰富关系数据的场景

✓ 直观的交互式建图，支持多源数据的导入
✓ 支持无向图、异构图、属性图，支持图的Schema设计，方便灵活的图属性查询、属性过滤
✓ 内置丰富的图查询、图挖掘算子库
✓ Python简易算法开发
✓ 接口灵活可定制，与SQL无缝对接
✓ 强大的图可视化展示
✓ 高速查询分析

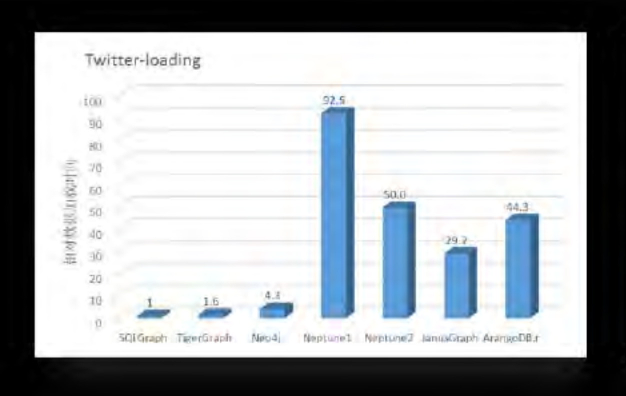


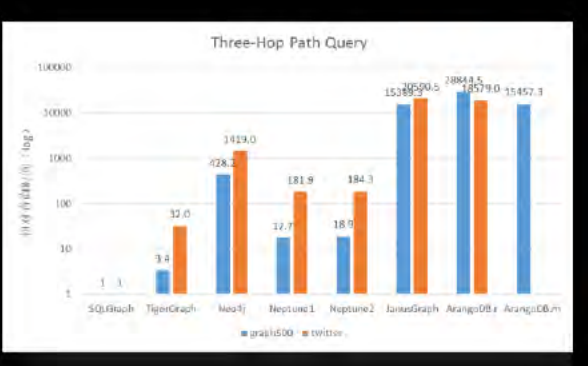
图计算引擎（SQLGraph）

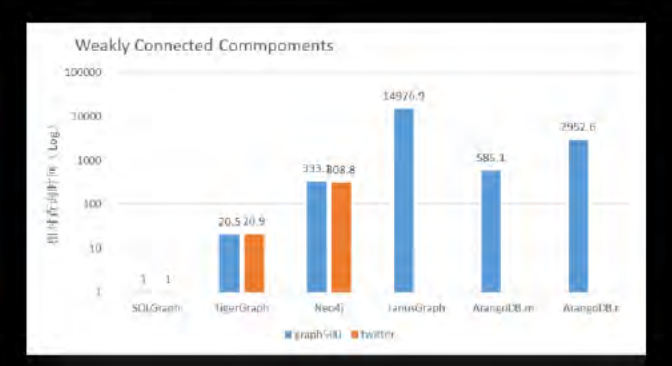


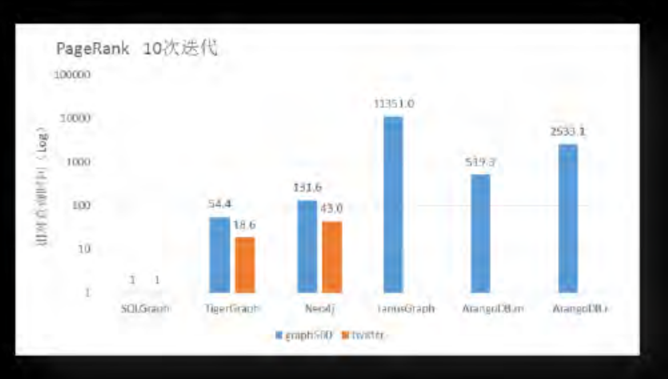
- 超高性能的关联查询分析
  - 支持**亿级节点**、**十亿级边**图计算的**秒级**响应
  - **亿级节点、边**导入图**十秒级**响应
- 国际主流产品对比
  - 与Neo4j、GraphX、TigerGraph等相比，图计算性能比快**10到1000倍**











内存数据库（IMDB）



产品定位

层级							
应用	嵌入式应用软件	桌面端应用软件	服务器端应用软件	移动端APP			
中间件	CORBA	DDS	J2EE	J2ME	图形库	...	
数据库	实时内存数据库：标准版、混合版、服务器版；SQL组件、高可用组件、集群组件；事务日志组件、Web服务组件、列数据库	关系数据库：Oracle、MySQL、PostgreSQL；达梦、金仓、神通、...					
操作系统	VxWorks	Linux	Windows	Android	iOS	...	
	锐华	道	天熠	神软	麒麟	...	
硬件平台	X86	ARM	MIPS	PPC	SPARC	Alpha	...
	龙芯	飞腾	申威	海光	兆芯	海思	...

开发环境：Qt IDE、GNU工具链、风河IDE、微软IDE、锐华IDE、道IDE、...

内存数据库（IMDB）



体系结构

层						
应用层	嵌入式应用软件	桌面端应用软件	服务器端应用软件	移动端APP		
接口层	C/C++	SQL	Java	Python	C#	...
扩展层	SQL组件	高可用组件	集群组件	事务日志组件	Web服务组件	列数据库
核心层	数据组织：数据布局管理、数据模型、数据字典	索引管理：Hash、Tree、自定义索引	事务管理：ACID、并发控制、恢复机制	运行时环境	安全：CRC32、RC4、Open SSL	
基础层	页管理	堆管理	数据库内存池（RAM或共享内存）			
适配层	龙芯+中标麒麟	飞腾+银河麒麟	X86+Linux	...		

内存数据库（IMDB）



服务体系

解决方案
- 全国产化实时数据管理解决方案
- 物联网设备实时数据管理解决方案
- 行业大数据实时数据管理解决方案
- 边缘计算实时数据管理解决方案

应用：嵌入式应用软件、桌面端应用软件、服务器端应用软件、移动端APP

中间件：CORBA、DDS、J2EE、J2ME、图形库、...

数据库
- 实时内存数据库：标准版、混合版、服务器版、SQL组件、高可用组件、集群组件、事务日志组件、Web服务组件、列数据库
- 关系数据库：Oracle、MySQL、Informix、达梦、金仓、神通、...

操作系统：VxWorks、Linux、Windows、Android、iOS、...、锐华、道、天熠、神软、麒麟、...

硬件平台：X86、ARM、MIPS、PPC、SPARC、Alpha、...、龙芯、飞腾、申威、海光、兆芯、海思、...

开发环境：Qt IDE、GNU工具链、风河IDE、微软IDE、锐华IDE、道IDE、...

技术服务：国产平台适配验证、基准测试/POC、数据库架构/设计、数据库性能优化、应用软件定制开发、第三方数据库迁移、第三方产品集成、...



应用组件产品-统一权限管理



统一权限平台系统架构

外部应用
应用系统A
应用系统B
单点登录
认证/鉴权服务
统一权限管理平台
应用管理
用户管理
角色管理
机构管理
资源管理
日志管理
REST API
日志系统
数据服务
应用数据服务
用户数据服务
角色数据服务
机构数据服务
资源数据服务
日志数据服务
数据存储
应用数据
用户数据
角色数据
机构数据
资源数据
日志数据

应用组件产品-可视化工具



基本架构

访问层
PC端
移动端
大屏端
其他终端
统一用户登录

业务层
智能分析
实时大屏
应用统计
数据资产
智能运营

功能层
预警层：事件预警
展现层
交互层：自助分析、即席分析、图文报告、中国式报表、移动交互、移动交互
分析计算：基本指标
占比、计数、平均、最大值、最小值、累计
时间累计占比、几何平均数、调和平均数、加权平均数、中位数、平方和
标准差、方差、偏度、峰度、增长状态、增长趋势
加工建模：数据准备、数据聚合、数据清洗、数据加速、数据建模
数据层：hadoop、spark、impala、hbase、实时数据、database、数据仓库、其他
系统管理
资源管理
组管理

支撑层
Nginx
uWSGI
url
app view logic
app view logic
app view logic
Database
app
view
logic

患者平均单个上皮细胞大小对比

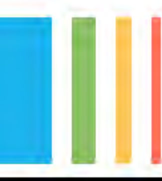
```
select case "Class"
  WHEN '0' THEN '普通患者'
  WHEN '1' THEN '乳腺癌患者'
  END AS "患者类别", abc."平均单个上皮细胞大小"
from
  (select "Class", avg("Single Epithelial Cell Size") as "平均单个上皮细胞大小" from keshcc_1574922222096 group by "Class") abc
```

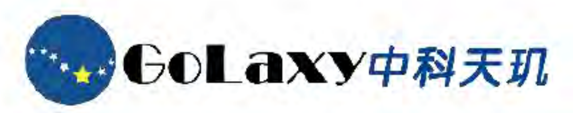
患者类别	平均单个上皮细胞大小
普通患者	2.12
乳腺癌患者	5.30

2 rows 0 seconds 运行

test

下载为 CSV 文件
下载为 Excel 文件
查看分析查询
组件配置定义
移出仪表板





行业应用产品/解决方案

应用体系



网信、宣传行业解决方案

舆情监测分析系统

安全行业解决方案

大魔镜系统

金融行业解决方案

金融舆情系统

军工行业解决方案

开源情报监测分析系统



舆情应用

情报应用

GoIN智能情报分析系统

核心技术与智能算法

自然语言处理 NLU/NLP	开放知识系统 OKS	图计算引擎 SQLGraph	内存数据库	……

天玑数据大脑

数据采集服务平台（WDE）	大数据存储管理平台（BDE）	可视化智能机器学习系统（BDA）

GoIN智能情报分析系统



以情报应用为主要抓手，通过对覆盖政治、军事、经济、社会等方面的大数据关联融合，构建贴近业务场景的分析挖掘、预测研判和智能生成系列工具，形成一套人脑决策和计算机智能共生的大数据分析环境及工具系统，辅助用户从大数据中萃取重大战略情报，并进一步转化为可操作的决策指挥能力

获情 Observe

用情 Act

OODA Loop

处情 Orient

出情 Decide

人机协同情报分析能力 GoIN GoLaxy Intelligence

深度理解能力 认知套件 CoreNLU CoreIMG DeepAccess

关联融合能力 知识套件 OKS OES

智能赋能能力 分析套件 SQLGraph AdaxNet BDA

大数据获取能力 XWDE

大数据管理能力 BDE

GoIN智能情报分析系统





交互式情报分析引擎

User Interface

过滤 透视 分析 探索

情报大数据融合分析引擎

内容分析（GA）	关联分析（NA）	时空分析（GA）
机器翻译 关键词 倾向性 文本摘要	节点重要度 社区发现 路径查找 子图挖掘	热力分析 区域探索 时间分析 轨迹分析

技术架构：基于动态知识结构的数据融合技术　深度内容理解与解析技术　深度信息检索与获取技术

大数据弹性一体化存储管理平台

Unified Storage APIs

MongoDB　SQLGraph　Elastic Search

Relational Database　Unstructured Data　Graphs　GIS　Files

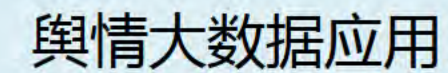
数据接入管理平台

Unified Accessing APIs

元搜索　社交媒体　文件上传　配置采集

舆情大数据分析


GoLaxy中科天玑


舆情大数据应用

中央网信办
地方网信办
宣传部
金融监管
网信整体解决方案
主被动结合的舆情监测系统
传播影响力分析系统
舆情报告
金融舆情监测系统
GoIN智能情报分析系统

网信：整体解决方案（地方网信办）



1个体系
- 技术标准规范

2个平台
- 大数据平台
- 指挥平台

2大业务
- 互联网信息监测
- 互联网宣传引导

层级	内容
解决方案	省网信平台；地市网信服务；政府舆情服务
业务分析	舆情监测；属地化管理；新闻传播效果监测系统；取证；网评；指挥通讯
数据处理	网信知识库；网信大数据平台
数据获取	数据采集



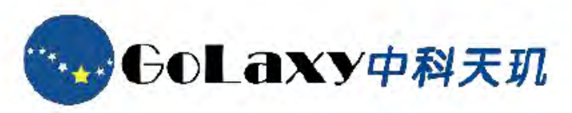
网信：行业大数据平台



- 以通用大数据平台+网信行业知识库，打造网信行业大数据平台，面向行业输出精准数据，面向客户输出定制化数据处理、基础性支撑平台能力

产品特色	关键指标	目标市场	对标竞品
• 行业知识闭环能力 • 行业数据分发能力	• 数据源覆盖及规模 • 行业知识库数量与精细度 • 数据标签准确性 • 行业知识来源	• 定制平台：部委/省级用户 • SAAS：基础性战略型产品	• 邦富/TRS底层数据标签

网信：行业大数据平台



以**精准数据、爬取数据、推送数据**为依托，通过技术手段整合多种数据源，建设**统一网信高性能大数据平台**。实现**多源异构数据融合处理、分层分类存储和基于网信业务场景的分析计算**。



集成数据采集、缓存、分发、存储、查询、检索、统计、分析、挖掘、可视化等海量数据全生命周期管理功能

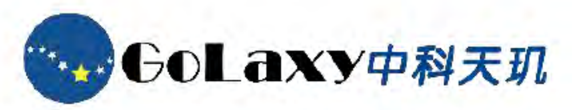








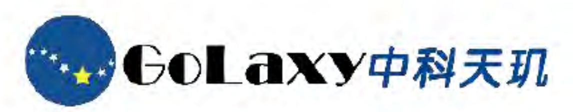
网信：指挥通讯系统



- 面向网信多级应急指挥协调的业务总线，实现省级网信办纵向指挥、横向协调的平台，通过业务系统对接，实现从信息发现、受理、研判、处置、落实、评估的业务闭环流程
- 要继续升级，充实省级网信办核心业务

产品特色	关键指标	目标市场	对标竞品
• 三级互通兼容 • 业务系统栈兼容（举报、舆情、传播监测与评估等）	• 上级指挥系统兼容性 • 业务系统兼容性 • 业务流程灵活性（权限、工作流可配置）	• 省级网信办	• 太极公司的指挥系统

网信：指挥通讯系统



- 形成上联中央、全省一体的指挥通讯，具备通信指挥、指令流转、指令效果评估等功能
- 实时安全、全面可靠、规范化、高可用性的中央、省级、地市三级指挥调度通讯

值班局长
带班领导
值班员
中央网信办
指挥中心
黑龙江
内蒙古
省级网信办
河南
江苏
西藏
四川
湖北
地市网信办
福建
云南
广东
管控指令
即时通信

- 遵照中央网信办指挥通讯系统建设规范，稳定、可靠、安全的运行
- 包含指挥通讯系统专网端和互联网端，采用双网协同方式，保障指令传输安全可控和数据存储安全可靠



网信：舆情监测分析系统



- 面向全通道舆情信息发现、预警、分析、追踪、出报的综合性舆情监测分析平台

产品特色	关键指标	目标市场	对标竞品
• 全通道数据融合 • 多角度舆情追踪（内容/目标/智能） • 精细化、智能化线索发现	• 通道数量及融合程度（全网溯源准确性、专题分析维度） • 舆情线索发现及时性、精准性	• 定制系统：部委/省级单位 • SAAS系统：地市单位	• TRS、云润（邦富）、智慧星光、人民舆情 • 微舆情、鹰击/鹰眼



网信：舆情监测分析系统

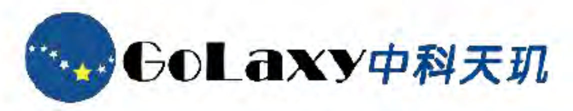


舆情系统是一款针对**舆情发现、分析研判、处置流转全生命周期管理**的数据分析服务平台。实现对**电子报、新闻、论坛、博客、微博、微信、APP、知乎、twitter、facebook、网民评论**等境内外全媒体数据的7*24小时实时采集、抽取、挖掘和分析处理，提供**精准舆情搜索、热点舆情聚合、负面舆情巡查、专项舆情监测分析、高危/高价值舆情预警、智能舆情报告**等服务。







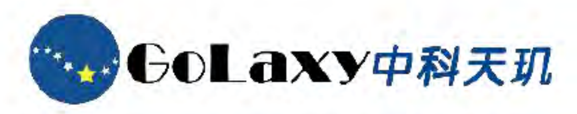


- 面向属地监管单位的属地信息媒介监管产品

产品特色	关键指标	目标市场	对标竞品
• 属地化：范围、形态 • 智能化：智能违规判断	• 属地信息媒介库规模与精准性 • 数据全面性：属地信息媒介数据获取召回率 • 管理维度：内容合法性、资质合法性、媒介影响力 • 识别精准性：非法媒介、非法内容自动识别精准性	• 省级/地市管理单位	• 天润嘉能属地管理



网信：传播监测与效果评估系统



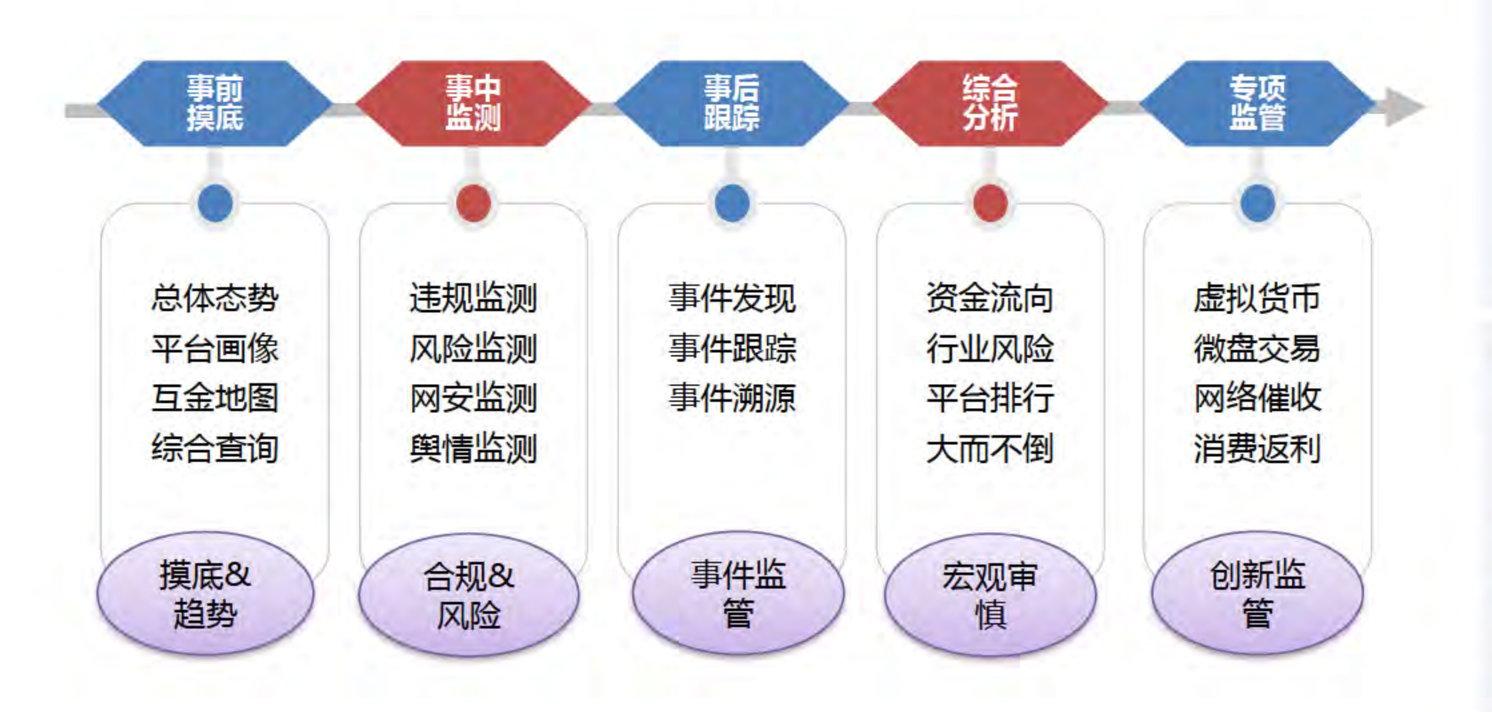
- 面向宣传调控业务的调控方案管理、核查、评估综合系统

产品特色	关键指标	目标市场	对标竞品
• 精细化核查：属地精细化、核查维度精细化 • 综合性评估：影响力评估、绩效评估 • 与指挥系统衔接，快速部署	• 核查维度：范围、位置、时间、过程 • 评估可信度：指标体系、数据、可视化 • 定制化速度：核查范围、核查维度、评估维度	• 省级网信办 • 媒体发达的地市	• 人民网新闻指挥调度系统

互联网金融监管

GoLaxy中科天玑

- 技术平台实时汇聚、分析来自于工商、法院、税务、互联网等多渠道数据，支持多维度、多业态、多场景的对互联网金融平台进行摸底发现、合规判断、风险识别、事件追踪等。
  - 全量工商、税务、法院、ICP备案库数据
  - 亿级企业信息、5000万+高管信息
  - 金融舆情：10万+版块，每天2000万+记录、30G+数据量





平台建设-技术体系



中科天玑联合中科院计算所在“互联网金融风险分析技术平台”的建设过程中承担了大量课题研究、算法建模以及系统开发工作：

知识库构建
金融知识库 互联网知识库
企业信息库 关联关系库
人员信息库 属性标签库

自然语言处理
特征词、关键词、实体词、信息提取、语义理解 ...

多源数据采集
网站、论坛、APP
微博、微信、FB ...

应用建模
业态类型识别
深度关联分析
信息溯源
文本情感分析
事件识别与预警

可视化展示
直观、生动、可交互 ...

互联网金融平台发现与精细化分类
互金舆情监测与舆情溯源
互金企业画像及关联图谱分析
互金网络广告监测
违法违规异常监测
疑似非法集资监测及立案分析系统
企业综合风险分析与预警
资本市场舆情监测系统

事前摸底
事中监测
事后跟踪
综合分析

平台建设-功能体系



事前摸底	事中监测	事后跟踪	综合分析	专项监管
总体态势 平台画像 互金地图 综合查询	违规监测 风险监测 网安监测 舆情监测	事件发现 事件跟踪 事件溯源	资金流向 行业风险 平台排行 大而不倒	虚拟货币 微盘交易 网络催收 消费返利
摸底&趋势	合规&风险	事件监管	宏观审慎	创新监管

教育产品线



教育科研大数据平台

各类高校的大数据相关专业

自2016年3月教育部批准设立大数据相关专业后，全国高等院校2956所，其中488所院校申请了大数据相关专业，未来还会有更多的院校申请建立这个学科。

使用大数据工具的学科实验室

在生物、医药、材料、金融、化学等需要大量统计分析数据的专业领域实验室，都越来越重视对大数据工具的掌握。

网络安全专业实验室

旨在为学校网络安全专业、社会网络安全评估机构、网络攻防测试机构、提供最新最全面的网络攻击手段以及防御能力评估。



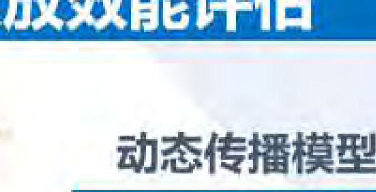
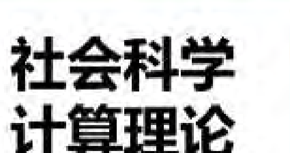


感谢聆听

应用场景：政治干预行动路径





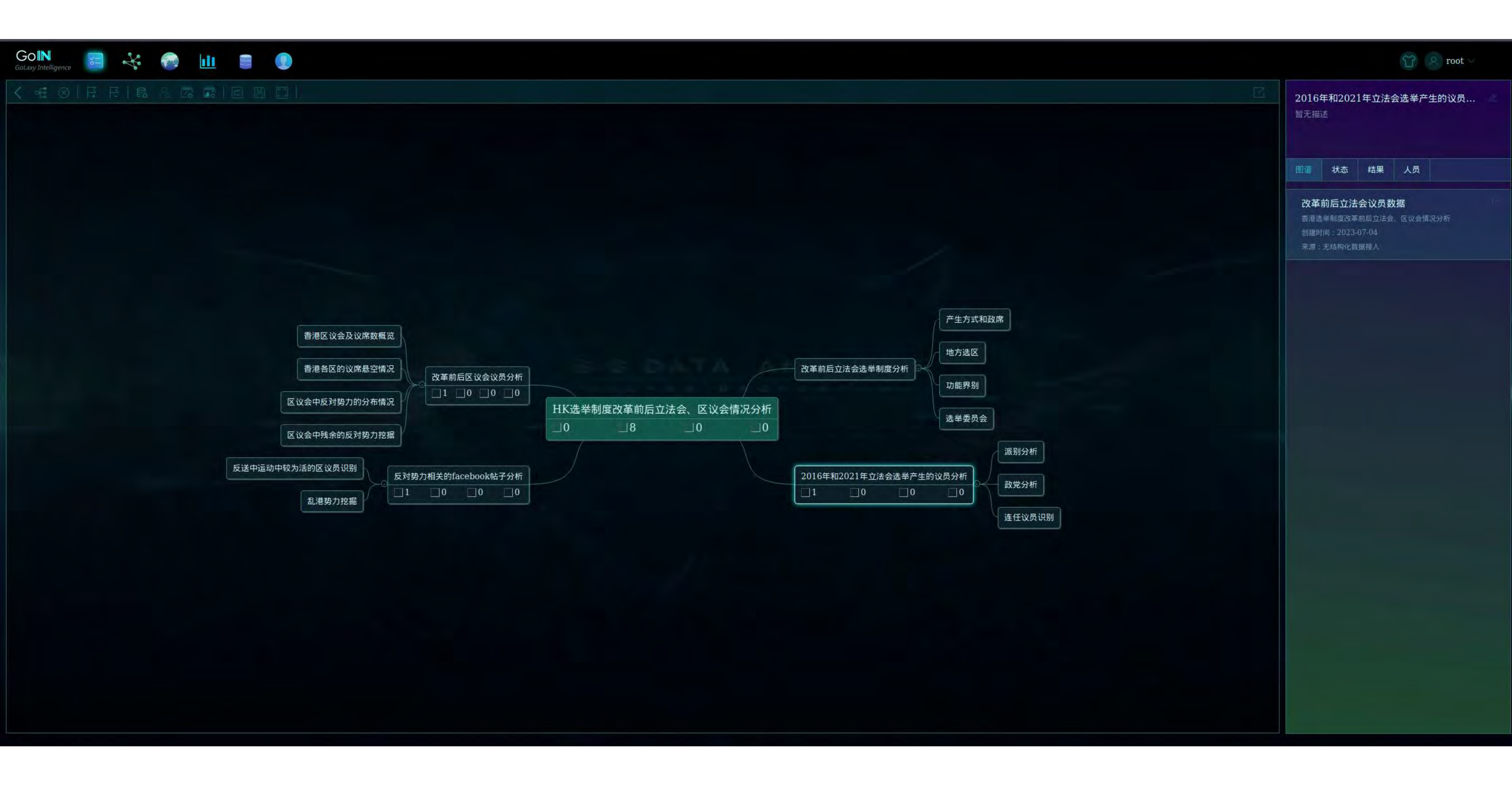




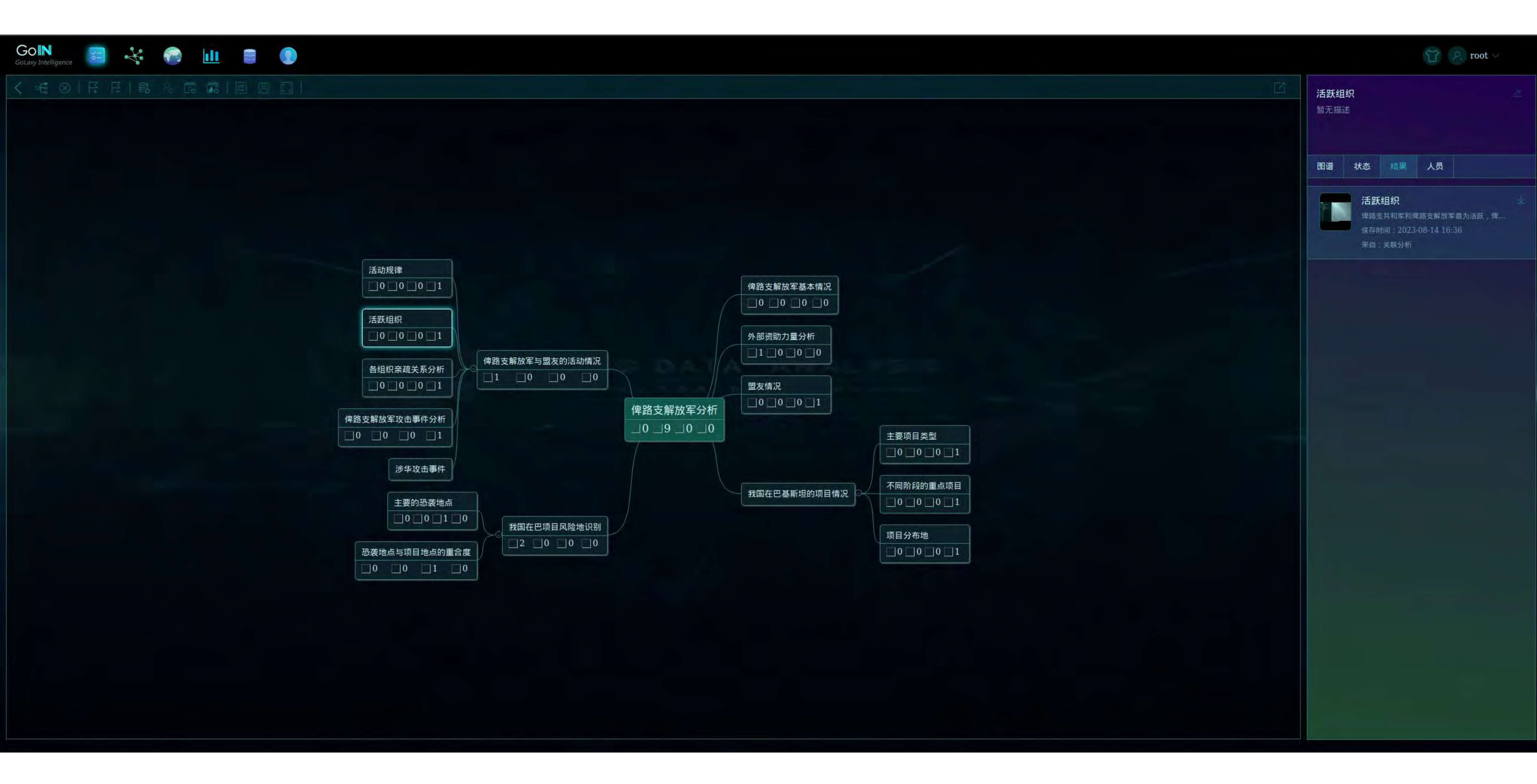
社会科学计算理论

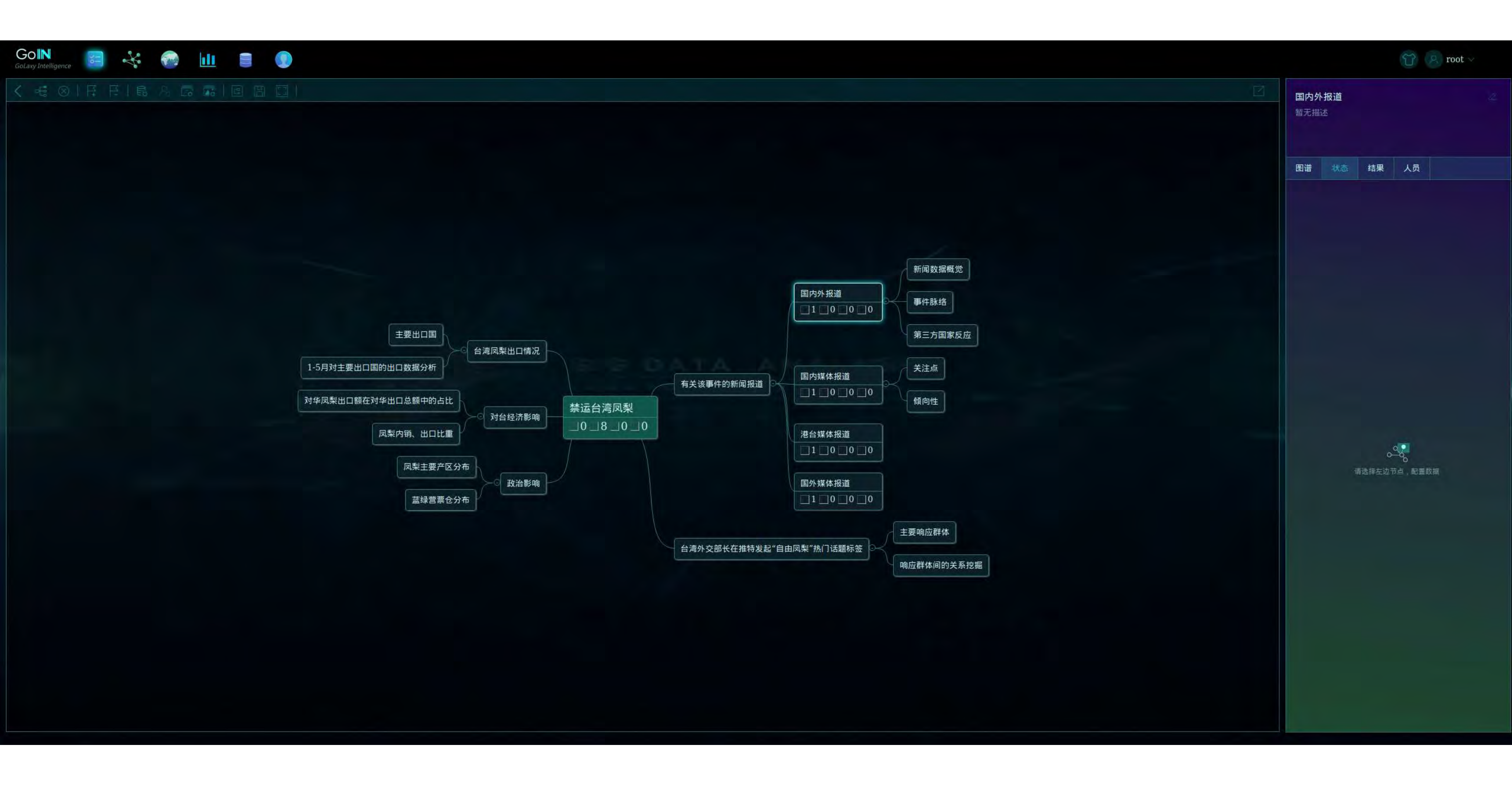
基于选举舆论争夺的政治过程干预模型

基于3.5%法则的舆论引导政治干预模型


GoIN
GoLaxy Intelligence
root
WestExec Advisors公司
本案例基于WestExec Advisors公司的官网数据及相关新闻，利用GoIN的智能分析和出报功能，尝试对该组织进行相对全面、深入的剖析，并生成报告。本案例除提...
图谱
状态
结果
人员
请选择左边节点，配置数据
WestExec Advisors公司
0 9 1 0
人员情况
1 0 1 0
人员概览
部分人员详情
美国前国防部是部长Michele Floumoy
Sergio Aguirre
Nitin Chadd
人员、组织间的利益关系
基本情况
1 0 0 0
合作伙伴
1 0 0 0
客户分析
盈利模式研判


GoIN
GoLaxy Intelligence
root
2016年和2021年立法会选举产生的议员...
暂无描述
图谱
状态
结果
人员
改革前后立法会议员数据
香港选举制度改革前后立法会、区议会情况分析
创建时间：2023-07-04
来源：无结构化数据接入
HK选举制度改革前后立法会、区议会情况分析
0 8 0 0
改革前后区议会议员分析
1 0 0 0
香港区议会及议席数概览
香港各区的议席悬空情况
区议会中反对势力的分布情况
区议会中残余的反对势力挖掘
反对势力相关的facebook帖子分析
1 0 0 0
反送中运动中较为活的区议员识别
乱港势力挖掘
改革前后立法会选举制度分析
产生方式和政席
地方选区
功能界别
选举委员会
2016年和2021年立法会选举产生的议员分析
1 0 0 0
派别分析
政党分析
连任议员识别


GoIN
GoLaxy Intelligence
root
俾路支解放军分析
俾路支解放军与盟友的活动情况
活动规律
活跃组织
各组织亲疏关系分析
俾路支解放军攻击事件分析
涉华攻击事件
我国在巴项目风险地识别
主要的恐袭地点
恐袭地点与项目地点的重合度
俾路支解放军基本情况
外部资助力量分析
盟友情况
我国在巴基斯坦的项目情况
主要项目类型
不同阶段的重点项目
项目分布地
活跃组织
暂无描述
图谱
状态
结果
人员
活跃组织
保存时间：2023-08-14 16:36
来自：关联分析


GoIN
GoLaxy Intelligence
root
禁运台湾凤梨
台湾凤梨出口情况
主要出口国
1-5月对主要出口国的出口数据分析
对台经济影响
对华凤梨出口额在对华出口总额中的占比
凤梨内销、出口比重
政治影响
凤梨主要产区分布
蓝绿营票仓分布
有关该事件的新闻报道
国内外报道
新闻数据概览
事件脉络
第三方国家反应
国内媒体报道
关注点
倾向性
港台媒体报道
国外媒体报道
台湾外交部长在推特发起"自由凤梨"热门话题标签
主要响应群体
响应群体间的关系挖掘
国内外报道
暂无描述
图谱
状态
结果
人员
请选择左边节点，配置数据

2021年度部门任务目标及分解

序号	部门	指标名称		责任人	年度目标（total）	Q1	Q2	Q3	Q4
1		经营业绩指标 (万元)	1.1新签有效合同						
			1.2净收入						
			1.3产品收入						
			1.4业务毛利						
			1.5行业利润						
2		重点工作任务项1:XXX	2.1xxx						
			2.2xxx						
			...						
3		重点工作任务项2:XXX	3.1						
			3.2						
			...						
4		......	4.1						
			4.2						
			...						

2020年度销售任务汇总表

单位：万元

部门	姓名	年度总任务（total）		Q1		Q2		Q3		Q4	
安全	孔令茹	1,000.00	800.00	150.00	120.00	250.00	200.00	300.00	240.00	300.00	240.00
安全	侯阿奇	1,000.00	800.00	150.00	120.00	250.00	200.00	300.00	240.00	300.00	240.00
安全	待招1	1,000.00	800.00	150.00	120.00	250.00	200.00	300.00	240.00	300.00	240.00
安全	待招2	600.00	480.00	90.00	72.00	150.00	120.00	180.00	144.00	180.00	144.00
安全	待招3	600.00	480.00	90.00	72.00	150.00	120.00	180.00	144.00	180.00	144.00
合计		**4,200.00**	**3,360.00**	630.00		1,050.00		1,260.00			1,008.00

行业事业部	行业	交付部门-待组织	甲方名称-需要	最终用户-需	最终用户所	项目名称	项目类型	项目阶段	项目金额	商机赢取率	销售人员	售前工程师	涉及自有产品情况-见表内信	外协部分说明（外	外部投入情况说明-计算所	预计签单日期	签单季度预估	项目联系人及联系	其他情况备注
安全行业事业部	安全	政法BG交付部	北京一三一八单位	1J	北京	大数据平台	产品销售	7-销售完成	3950000	100%	孔令茹	于明远	BDE\数据治理平台、知识图谱	技术服务	无	30/03/2021	2021年第一季度		
安全行业事业部	安全	政法BG交付部	计算所	1J	北京	综合业务平台	产品销售+开发	5-招标投标	6480000	80%	孔令茹	于明远	BDE\数据治理平台、知识图谱	OCR软件识别	无	30/06/2021	2021年第二季度		
安全行业事业部	安全	政法BG交付部	计算所	14J	北京	开源数据采集子系统	产品销售+开发	3-需求确认	1260000	80%	孔令茹	于明远	数据采集		无	30/06/2021	2021年第二季度		
安全行业事业部	安全	政法BG交付部	计算所	11J	北京	应用系统软件开发	产品销售+开发	5-招标投标	3636000	＜50%	孔令茹	于明远	数据采集、舆情产品		无	30/09/2021	2021年第三季度		
安全行业事业部	安全	政法BG交付部	计算所	11J	北京	人工技术服务	产品销售+开发	2-商机确认	2790000	＜50%	孔令茹	于明远	舆情产品		无	30/09/2021	2021年第三季度		
安全行业事业部	安全	政法BG交付部	辽宁T	辽宁T	沈阳	QB分析系统	产品销售+开发	1-商机发现	3000000	＜50%	孔令茹	于明远/孙涛	GOIN		无	30/09/2021	2021年第三季度		
安全行业事业部	安全	政法BG交付部	河南T	河南T	河南	T公司人员数据调研	数据服务	4-方案设计	50000	80%	孔令茹	刘誉飞	数据调研		无	30/08/2021	2021年第三季度		
安全行业事业部	安全	政法BG交付部	国测中心	国测中心	北京	APY知识库	产品销售+开发	4-方案设计	1000000	＜50%	孔令茹	孙涛	数据采集		无	30/10/2021	2021年第四季度		
安全行业事业部	安全	政法BG交付部	河南T	河南T	河南	YH信息	集成	7-销售完成	1700000	100%	孔令茹	于明远	设备		无	30/01/2021	2021年第一季度		

			待组	需要	需	所								见表内信	(外	计算所			联系		反映
渠道	安全事业部	安全	政法BG交付部	深圳局	深圳局	深圳	数据专题分析	产品销售+开发	3-需求确认	3000000	50%	侯阿奇	于明远	无	无	无	30/11/2021	2021年第三季度			基本 部里 需要 做数 前没
渠道	安全事业部	安全	政法BG交付部	广东厅	广东厅	广东	小魔镜3.0	产品销售	1-商机发现	300000	70%	侯阿奇	于明远	暂无	无	无	30/09/2021	2021年第三季度			客户 造， 求更 维护 改造 务器
渠道	安全事业部	安全	政法BG交付部	四川厅	四川厅	四川	数据治理	产品销售+开发	3-需求确认	4800000	50%	侯阿奇	于明远	暂无	服务器	无	25/11/2021	2021年第三季度			预计 标流 委会
渠道	安全事业部	安全	政法BG交付部	福州局	福州局	福建	桃T	数据分析服务	4-方案设计	60000	80%	侯阿奇	孙涛	暂无	无	无	25/03/2021	2021年第一季度			福州 ，完 交付
渠道	安全事业部	安全	政法BG交付部	甘肃省GA办	甘肃省GA办	甘肃	舆情态势感知	产品销售+开发	2-商机确认	800000	50%	侯阿奇	于明远	暂无	无	无	26/12/2021	2021年第四季度			上报 质， 办公 以后
渠道	安全事业部	安全	政法BG交付部	西藏AQ厅	西藏AQ厅	西藏	邮件分析系统	产品销售+开发	2-商机确认	3000000	60%	侯阿奇	于明远	有	服务器	无	27/09/2021	2021年第三季度			现在 及到 系统 在真
渠道	安全事业部	安全	政法BG交付部	四川省AQ厅	四川省AQ厅	四川	数据治理	产品销售+开发	1-商机发现	500000	<50%	侯阿奇	于明远	暂无	无	无	25/11/2021	2021年第四季度			针对 增值 价50
渠道	安全事业部	安全	政法BG交付部	青海省公安厅	青海省公安厅	青海	治安一体化防控	产品销售+开发	2-商机确认	10000000	<50%	侯阿奇	于明远/孙涛	暂无	无	无	25/11/2021	2021年第四季度			九大 体规 案， 性工 评估 度。

	关键举措 (我们打算如何实现)	评价标准 (对E列目标的达成结果定义,	完成时间 (对应E列)	负责人	Q1达成目标	Q1评价标准	Q2达成目标	Q2评价标准	Q3达成目标
效签约额	*项目1-填写具体项目名称*		*格式：2020-12-31*	*填写项目主要责任人*	*填写具体金额*	*可衡量的评价标准*	*填写具体金额*	*可衡量的评价标准*	*填写具体金额*
	项目2								
	项目3								
收入	*项目1*								
	项目2								
	项目3								
品收入	*项目1*								
	项目2								
	项目3								
务毛利	*项目1*								
	项目2								
	项目3								
业利润	*项目1*								
	项目2								
	项目3								
	具体措施1				*请填写目标产出或效果*				
	具体措施2								
	具体措施3								
	具体措施1								
	具体措施2								
	具体措施3								

评标结果确认函

中化商务有限公司:

项目编号：0747-2060SCCZD820

项目名称：xx 实体抽取与检索技术（二次招标）

我单位已审阅以上项目的评标报告，同意以下评标结果：

中标候选人	投标总价（人民币元）	综合得分	排名
中科天玑数据科技股份有限公司	252,000.00	82.54	1
北京中科浩晖科技股份有限公司	258,000.00	68.77	2

确认以此对外发布评标结果公示，并在公示期结束后向排名第一的候选中标人发出中标通知书。

单位名称：中国人民解放军 31040 部队

盖章/签字：

日期：2020 年 11 月 9 日

版本：1.0

密级：内部

智能宣传系统（群体分析子系统）概要设计说明书

	智能宣传系统（群体分析子系统）概要设计	1.0	1	

文档历史记录

日期	名称	版本	状态	签名

备注：状态表示为：1.新建 2.完成未审 3.完成已审 4.完成归档

文档修改记录

[目的与内容] 记录本文档的修改历史，包括版本、日期、变动内容和修改类型。

编号	版本	描述 （包括：日期，变更的图、表、段落号，标题或简单描述）	修改类型 A，M，D
1	1.0	初次编写文档	A
2			
3			
4			
5			
6			

备注：修改类型表示为：　**A -** 增加　**M -** 修改　**D -** 删除

目录

目 录

1.引言

1.1.编写目的

该概要文档撰写的主要目的，是在需求设计的基础上，就该系统的接口、模块、库表、界面等几大部分进行深入分析和讨论并形成交互框架和视觉框架的过程，为用户研究和详细设计的无缝链接起到很好的桥梁作用，该文档预期的读者包括负责人员、开发人员、测试人员、运维人员，以及其他对该系统的开发和使用必须了解的人员。

1.2.背景

本概要设计所关联的系统是智能宣传系统，该任务是由智能宣传小组提出具体应用需求，并由王永庆老师负责需求、ui 组人员负责设计、组内开发人员负责开发、部署以及后期运维工作。

本系统为智能宣传的子系统，为辅助系统，属于智能宣传系统中的群体分析模块。本系统主要由刘金虎、史铂深二人负责。

1.3.名词解释

缩写、术语	解 释
社交网络（SNS）	Social Network Site，专指旨在帮助人们建立社会性网络的互联网应用服务。
JSON	Javascript Object Notation，一种轻量级的数据交换格式。
群体	不同个体按某种特征结合在一起，进行共同活动、相互交往，形成了群体，本系统中特指一系列有共同特征的社交网络帐号。
用户画像	本系统中指某个社交网络账号的用户的属性、行为与期待的数据转化集合。

2.任务概述

2.4.目标

本项目主要针对群体画像数据数据，实现群体分析功能。

2.5.开发与运行环境

2.5.1.开发环境的配置

类别	标准配置	最低配置
计算机硬件	16G 内存，4 核 2.2GHZ 以上，千兆网卡，500G	8G 内存，2 核 2.2GHZ 以上，千兆网卡，500G
软件	Windows、MAC os 、JDK、Tomcat、DT	Windows、JDK、Tomcat、DT
网络通信	TCP/IP，HTTP	TCP/IP，HTTP
其它	无	无

2.5.2.

2.5.2.运行环境的配置

类别	标准配置	最低配置
计算机硬件	64G 内存以上，8 核 2.2GHZ 以上，千兆网卡，1T	32G 内存，4 核 2.2GHZ 以上，千兆网卡，500G
软件	CentOS、JDK、Tomcat、DT	CentOS、JDK、Tomcat、DT
网络通信	TCP/IP，HTTP	TCP/IP，HTTP
其它	无	无

2.5.3.

2.5.3.测试环境的配置

类别	标准配置	最低配置
计算机硬件	16G 内存，4 核 2.2GHZ 以上，千兆网卡，500G	8G 内存，2 核 2.2GHZ 以上，千兆网卡，500G
软件	CentOS、JDK、Tomcat、DT	CentOS、JDK、Tomcat、DT
网络通信	TCP/IP，HTTP	TCP/IP，HTTP
其它	无	无

2.5.4.

2.3.需求概述

根据需求说明书，该系统的主要功能由以下几项组成：

1.群体创建

支持添加不同的分析选项，选择完成后向数据方发送群体分析创建的信息。通过 DT 异步返回群体数据。（分析选项是否持久化待定）

2、群体分析

群体分析模块分群体展示群体的相关数据，具体数据向数据方请求。

3、用户画像

单用户画像模块展示各个用户的相关数据，具体数据同样向数据放请求。

模块主要功能应用的场景以及输入和输出如下所示：

功能	场景	输入	输出
群体创建	根据关键信息创建一个新群体	分析选项：关键词，重点用户 分析起止时间 社交平台 地域	一个用户群体（通过 dt）
用户列表	展示用户列表		
用户画像	用户详情页		单个用户画像分析
群体画像	群体详情页		单个群体画像分析

2.4.条件与限制

条件如下：

条件 1：模块所需要的数据包括：用户画像分析，用户详细数据 由实验室数据平台提供

条件 2：模块所需要的生产环境由实验室提供

条件 3：模块所需开发环境由业务人员自行选择

条件 4：开发环境和测试环境由业务人员自行调度

条件 5：系统试用期间，中心的业务人员能够投入一定比例的时间试用该系统，并就该系统存在的功能和业务问题给出反馈；

条件 6：开发期间若有业务上的问题，业务人员能够参与一些必要的讨论，并就问题能够给出一些建议；

条件 7：界面设计由组内人员讨论，ui 组实现具体设计

限制 1：进度限制，按照规定的日期进行开发；

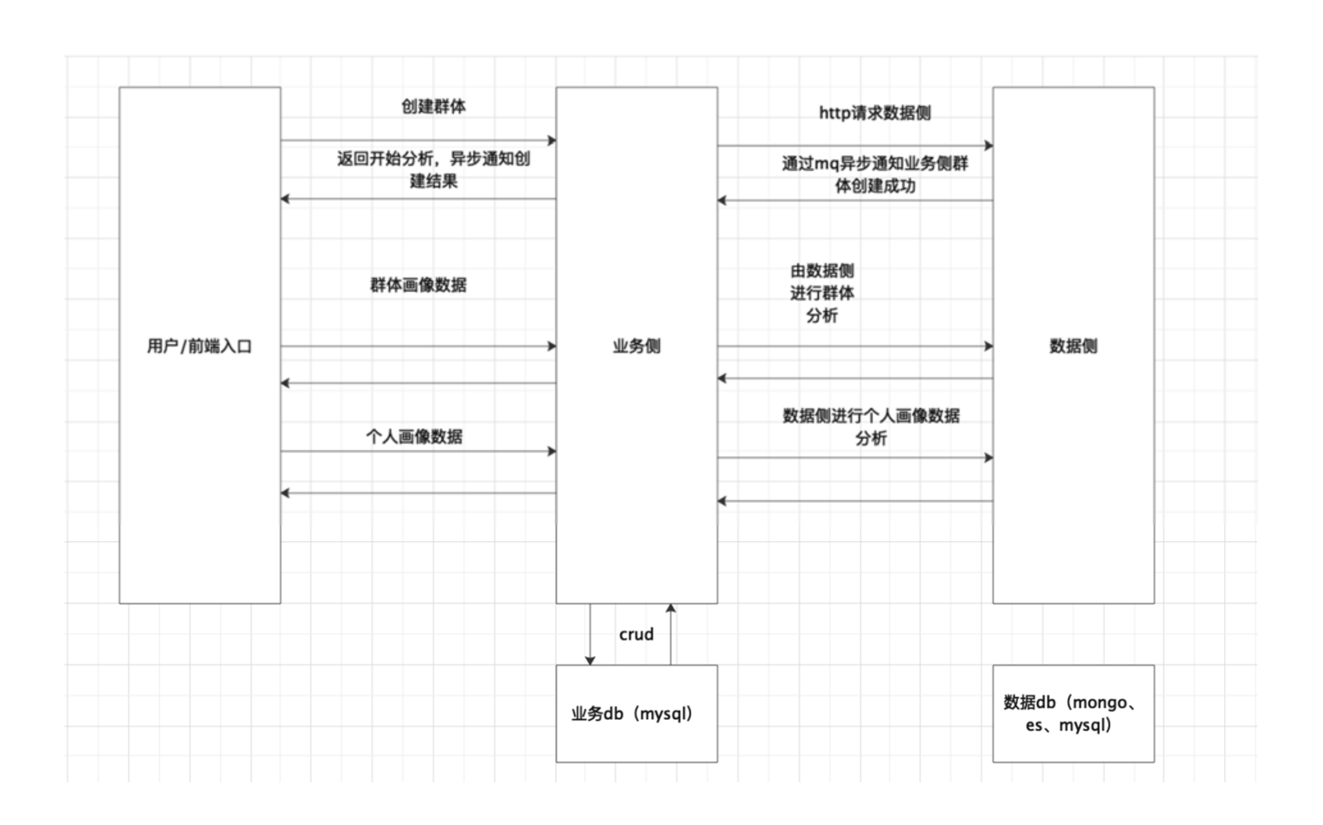
限制 2：功能限制，参考需求和时间人员的投入来确定范围；

限制 3：设备限制，系统架构的设计应在计算和服务资源的限制范围内进行；

3.总体设计

3.1.基本设计概念和处理流程

群体分析子系统的整体架构如下：



群体分析子处理流程如下：

1.由前端(用户)根据关键字开始群体分析，由业务侧告知数据侧

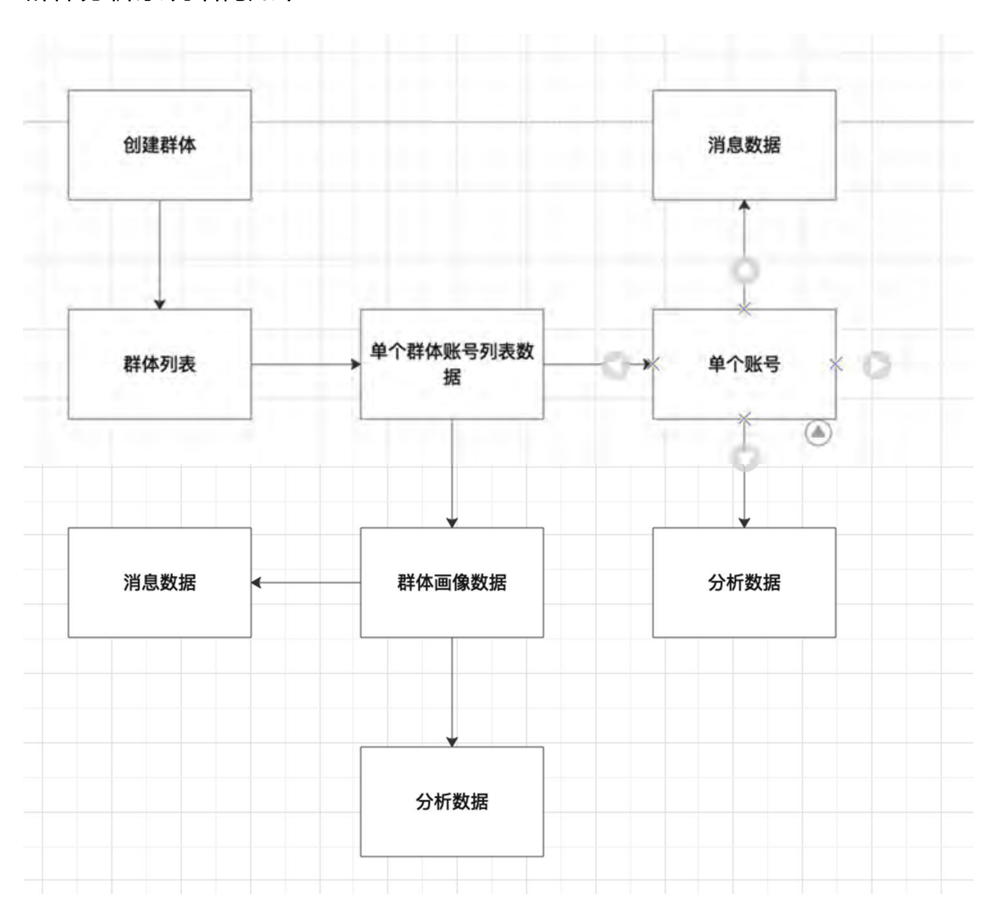
2.群体分析是一个由数据侧进行的耗时的操作，业务侧和数据侧通过 mq 异步通知，业务侧和前端可通过 socket（暂定）通知。

3.群体画像数据由前端请求，业务侧向数据侧请求数据并由数据响应时长等其他因素决定是否在业务 db 进行缓存。

4.个人画像数据由前端请求，业务侧向数据侧请求数据并由数据响应时长等其他因素决定是否在业务 db 进行缓存。

3.2.结构

群体分析系统结构如下：



说明：群体分析系统主要分为群体画像数据和个人画像数据，画像数据分为统计分析数据和消息数据。

3.3.其他需要说明的问题

无

4.接口设计

前端调用接口统一使用 restful api 操作，接口文档待与前端讨论后另出文件。

5.模块设计

5.1 群组账号列表

群体账号列表由数据侧在创建群体后将各个账号的基本信息返回，存入业务侧的数据库。业务侧维护账号和群体的映射关系。基本信息包括（粉丝数，关注数，消息数等）。

5.2 群体分析页面

群体分析页面的数据由数据侧提供，并根据响应时间等其他因素确定是否由业务侧进行缓存，并日更新。群体分析数据包括群体活跃度分析，关键词和话题等数据。

5.3 个人分析页面

个人分析页面包括基础数据和统计分析数据，基础数据由业务侧直接提供，统计数据则实时访问数据侧获取。个人基础数据包括粉丝数，关注数，发帖数等。统计数据包括，活跃度分析，关键词，话题等。并根据不同的目标账号类别进行话题粒度的区分，普通账号不进行话题粒度的区分，媒体账号和重点账号进行话题粒度的区分。

6.数据库设计

表名：target_account_basic_info

字段名	类型	说明
target_account_id	String	账号 id uid
user_name	String	用户显示名称
screen_name	String	@后的名称
real_name	String	真实姓名
gender	tinyint	性别-1 无效 0 女 1 男
lurl	String	头像地址
profile_url	String	头像地址
description	String	描述
create_time	Int	用户创建时间戳
gtime	Int	爬取时间戳
Is_verified	tinyint	是否认证 0 普通用户,1vip 用户 -1 未知
verify_info	String	认证信息
Is_real	tinyint	是否实名认证 (1:实名；0:非实名,-1:未知)
home_url	String	用户主页
pburl	String	横幅
banner_url	String	横幅
location	String	用户所在地
email	string	用户邮箱
birthday	String	用户生日
platform	String	平台
Update_time	Int	更新时间戳

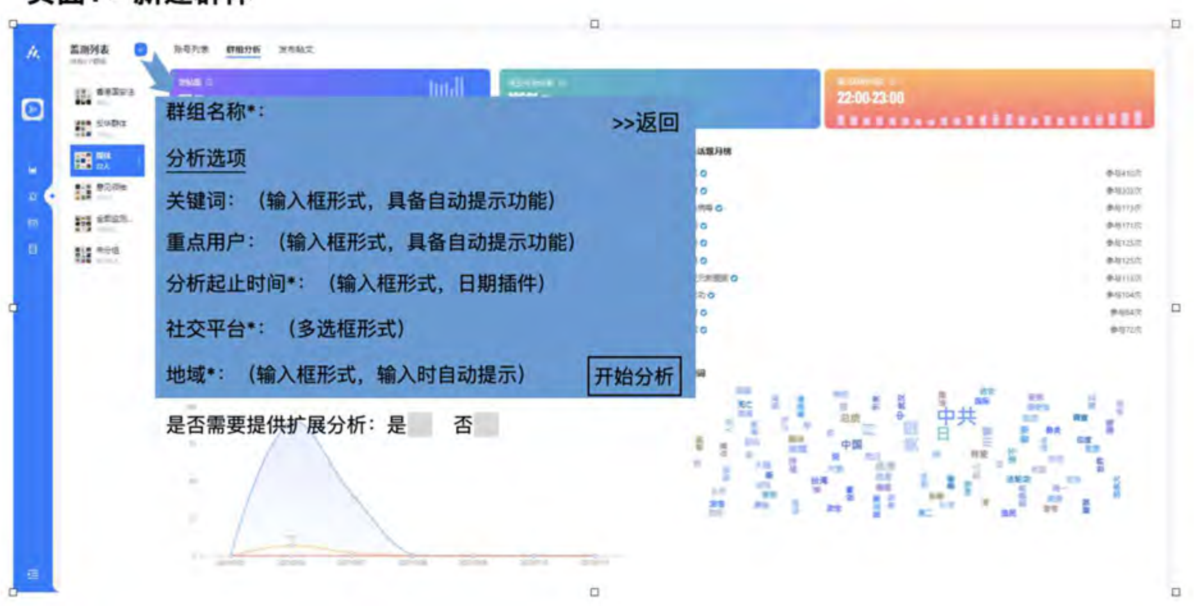
表名：target_account_statistic_info

名称	类型	说明
target_account_id	String	用户 id

follow_count	Int	关注数
fans_count	Int	粉丝数
status_count	Int	发文数
forward_count	Int	转发数
comment_count	Int	评论数
yes_follow_count	Int	昨天 关注数
yes_fans_count	Int	昨天 粉丝数
yes_status_count	Int	昨天 发文数
yes_forward_count	Int	昨天 转发数
yes_comment_count	Int	昨天 评论数

7.界面设计

页面1：新建群体



页面2：账号列表（此处功能同一期账号列表页）



页面3：群组分析（此处功能分块同一期群组分析页，具体信息展示参照下图）



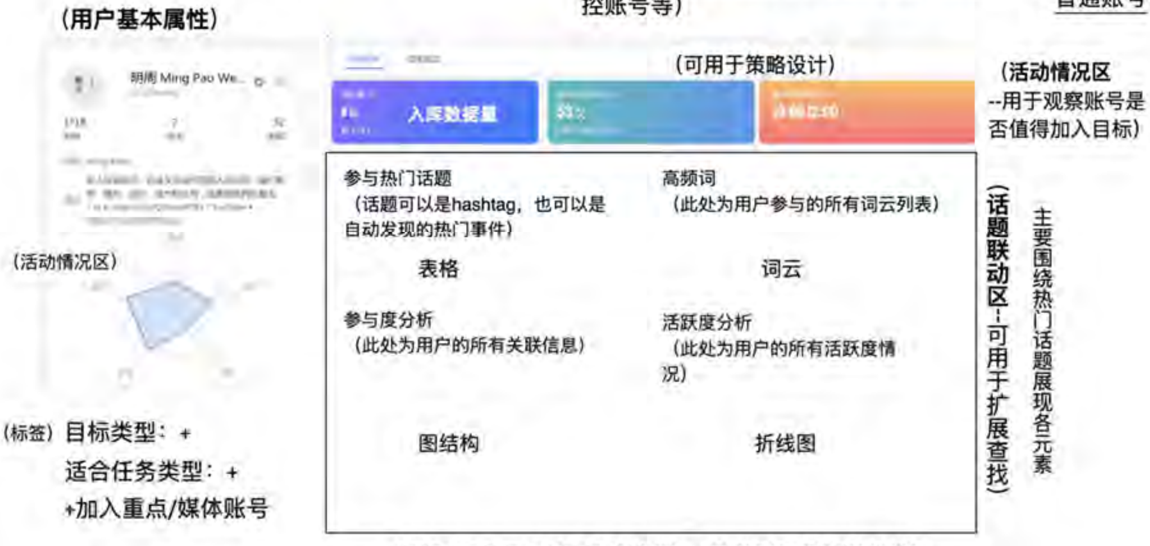
页面4：相关贴文（一期发布贴文页删除，改为点击用户高频词、热门话题后的相关贴文展示）



页面5-1：用户画像（点击账号链接进入）

设计思路：通过对当前人物的观察，将用户加入目标列表（不同账号的显示方式应有区分，例如：XX媒体、被控账号等）

普通账号



（普通用户仅展示全部的情况，不做话题级别的分析）

页面5-2：重点账号画像

设计思路：通过对当前人物的观察，将用户加入目标列表（不同账号的显示方式应有区分，例如：XX媒体、被控账号等）

重点账号




页面5-3：媒体账号画像
设计思路：通过对当前人物的观察，将用户加入目标列表（不同账号的显示方式应有区分，例如：XX媒体、被控账号等）
媒体账号
（用户基本属性）
明周 Ming Pao We...
（活动情况区）
（标签）目标类型：+
适合任务类型：+
（可用于策略设计）
入库数据量
（活动情况区
--用于观察账号是否值得加入目标）
参与热门话题
（话题可以是hashtag，也可以是自动发现的热门事件）
表格
参与度分析
（在未选择热门话题时，此处为用户的所有关联信息；选择话题后，显示与话题相关的用户关联情况--所有都为2跳以内的情况，并在高频词后加入对应话题提示）
图结构
高频词
（在未选择热门话题时，此处为用户参与的所有词云列表；选择话题后，显示与话题相关的词云情况，并在高频词后加入对应话题提示）
词云
活跃度分析
（在未选择热门话题时，此处为用户的所有活跃度情况；选择话题后，显示与话题相关的用户活跃度情况，并对比整体整体活跃度情况--双y轴展示，并在高频词后加入对应话题提示）
折线图
（话题联动区--可用于扩展查找）
主要围绕热门话题展现各元素

页面5-4：相关贴文列表


明周 Ming Pao We...
目标类型：+
适合任务类型：+
+加入重点/媒体账号
（可用于策略设计）
入库数据量
相关贴文（话题->关键词）
（点击高频词，展示用户于该高频词（及话题）下的相关贴文信息--普通用
Apple Daily HK 蘋果日報
Apple Daily HK 蘋果日報

8.安全保密设计

1.本系统所使用的安全模块为天玑公司提供，安全性由天玑公司保障。

2.任何账号由产品/项目经理所保管，不得外泄。

3.硬件服务器由实验室提供。

9.维护设计

1. 按照维护计划中要求，对本系统主要软件、硬件设备进行经常性地维护与监视，保证其正常工作。
2. 由专门维护人员进行周期性的维护，并如实记录维护报告与修理报告。
3. 按照维护计划中的要求，定期对系统设备进行检修、维护，保证系统满足运行要求。
4. 提交维护报告与修理报告，以便客户登录系统的日常工作状况，报告系统维护及修理情况。

编号：

密级

“××××工程”4月进展报告

1. 当前工作进展

围绕“战、建、备”协同推进的总体思想，4月份全面启动业务需求调研工作，并根据调研情况反馈同步推进工程建设工作。此外，由于工程行政总体变更为由正宫部主管，为配合工程总体调整，启动项目总体建设实施方案以及抓建模式调研工作，并初步形成总体论证调整思路。具体开展工作包括：

（1）业务需求调研。明确沈华伟、王永庆、王石、张瑾、邱强、李蕾为工程展场核心人员，持续赴福州 61716 不对开展业务需求调研。作为工程理论总体，在全面了解业务单位的工程建设目标的基础上，总结理论瓶颈、梳理理论体系，与业务单位就任务设置、执行路径、成果形式等展开深入讨论，初步确定基础理论创新群的 2023 年度任务清单。

（2）配合调研工作。为配合正宫部的调研活动工作开展，分别于福州、北京两地开展调研活动，就总体推进过程中如何理论先行牵引工程建设任务的方式及关键问题进行讨论。在计算所调研期

间，陈熙霖所长介绍所在认知方向上的战略布局，强调在工程建设上计算所不计成本不计投入优先保障。宣传局李军局长对计算所的技术积累及工作进展表示肯定。

（3）总体论证优化。调研活动结束后，正宫部着手组织总体论证优化工作。计算所委派沈华伟、张浩、王永庆、王石、张瑾、邱强加入论证组，完善指挥决策与基础理论部分的均势能力需求、建设目标、建设内容与任务安排。截至月底前，根据国家宗提展的任务要求已形成能力需求、建设目标和建设内容初稿，并对基础理论的抓建模式给出建设性意见。

（4）工程建设推进。跟随业务调研工作进展，计算所数据智能系统研究中心、网络数据科学与技术重点实验室、智能信息处理重点实验室、前瞻研究实验室 4 中心协调联动，于社交基期刃调度、指挥决策、内容智能生成、事件分析等方面，开展需求设计、技术方案设计和原型系统设计工作，相关方案已形成技术实施方案并入工程总体建设目标。

2. 下一阶段工作考虑

围绕持续推进的工程总体论证工作以及初步确定的理论研究体系和年度工作任务清单，下一阶段工作主要分以下几个部分开展：

（1）总体论证优化。持续跟进，对指挥决策、理论体系以及理论部分的抓建模式进行优化调整，以理论牵引为目标，进一步明确理论工作的 5 年期任务开展方式及实施路径。

（2）业务需求调研。在已初步明确的 2023 年任务清单基础上，进一步明确任务要求。相关已具备研究条件的课题内容，优先联合优势单位开展合作研究。对未具备研究条件的课题内容，调研工程需求，形成指南建议。

（3）工程建设推进。瞄准 7 月沿袭任务，组建优势力量，重点就数据采集、内容生成、社交基期刃调度等系统建设形成建设方案，全面启动建设任务。

脚本文字稿

一、开场白

非常感谢院领导莅临我所检阅和指导工作，我谨代表计算所网络数据科学与技术重点实验室向各位领导汇报“天玑智能宣传系统 GoPro”。

二、智能宣传研究背景

近年来，国际形势风云变幻，暗潮涌动，以美国为代表的西方国家频频借香港问题、新疆问题、新冠病毒、贸易战在网络空间向我发动攻势强大的舆论战，严重威胁我国意识形态安全和政权安全。面对新形势新任务，习近平总书记多次指出“要增强阵地意识，敢于发声、敢于亮剑，尽快掌握网络舆论场上的主动权，不能被边缘化了。”因此，创新利用新兴网络技术，研发先进的技术系统，改变我国在网络舆论斗争的“守势”和“弱势”现状，弥补网络信息流 “逆差”、形象 “反差”、实力“落差”，成为科研工作者需要解决的时代命题。

面向国家重大需求，我所将智能宣传列为所内 8 大优先支持发展方向，并在网数实验室成立专门的系统部，由 2 名正高、5 名副高等 50 余人组成的研究团队，在尚未有国家相关项目和经费的支持下，部署并开展学科方向论证、关键

技术攻关、演示验证系统研发等工作，下面向各位领导详细介绍一下智能宣传系统。

三、智能宣传系统介绍

系统着着眼于新时期网络舆论引导和认知对抗的迫切需求，创新利用人工智能、大数据、社会心理、信息传播等技术成果，研究网络舆情智能传播、引导与利用技术，研发面向境外非受控环境的智能宣传系统，具备如下四大能力：

一是特定区域政治态势的智能感知。围绕当前中美博弈、香港问题、新疆问题、台海问题等涉我重大议题，基于 WDE 平台广泛采集特定区域政治、外交、军事等方面的数据，通过自然语言处理技术，实时聚合出每日热门政治事件，并深度分析该事件的热度、影响力、相关话题、演变趋势，以及与该事件相关的主要媒体和跨平台社交账号的情感倾向、政治立场和观点看法。

二是境外目标对象的智能锁定。基于实验室积累十年的情报分析技术和 GoIN 系统，深入挖掘热门政治事件的幕后组织和关键人物。以近期香港国安法为例，通过广谱关联技术从监测的 18 万香港 Twitter 用户中，挖掘出与该事件相关的 3000 多名主要人物，100 多名网络意见领袖，并利用人物画像技术多维度刻画人物的行为特征、活动规律、政治倾向、观点立场等，为舆论引导提供精确的目标对象。

三是浸入式宣传内容的智能生成。根据特定区域的政治态势和受众对象的习惯爱好，充分利用人工智能技术辅助创作引导内容，如视频所示，基于联想功能可自动续写与输入内容相关的一段文本；相关素材推荐为创作者提供多源海量素材；衍生文案能基于原始文稿洗出若干篇语义不变、表述不同的新文稿；智能创作支持一体化的长文本内容生成。

四是非受控环境下的智能推送。境外社交媒体风控严格，在境外规模化、智能化、组织化培育社交机器人账号需要大量的资源投入和技术操作。目前本团队在 Twitter、Facebook、Telegram 等平台培育了数百个拟人化的机器人账号，并提供一站式跨社交媒体操作和调度平台，实现自动批量发帖、评论、转发，为大规模、精准、安全投放引导内容提供技术基本。

四、工作总结与展望

当前，我们仅是迈开万里长征的第一步，自主研发的智能宣传系统作为新兴事物，初步具备政治态势感知、群体心理画像、内容智能生成和信息智能投送能力，但离全面服务国家在境外网络舆论斗争的需求还有一定差距。

未来我们期盼在科学院领导和所领导的大力支持下，与中央网信办、中宣部、军科委、安全部等部门加强合作，明确任务需求，加大资源投入，打造出能真正自动讲好中国故

事、发出中国声音、扩大中国影响力的技术平台，为早日实现“东风压倒西风”提供全方位的技术支持。



QB 分析监测系统

需求说明书

中科天玑数据科技股份有限公司

2024 年 03 月 06 日

修改历史

时间	需求来源	版本	作者	修改内容
2024.03.06	项目客户	V1.0	杨菊海	新增

目录

1 简介

1.1项目背景介绍

各业务部门随着开展业务工作不断增多，掌握的数据种类不断丰富、数据结构不断异化、数据总量急速增长，导致业务人员缺少对数据的统一管理以及如何基于智能化引擎深挖数据价值，提升数据分析的智能化水平。因此亟需建设 QB 分析监测平台，具备“多源采集、专题监测、QB 分析”的能力，用来满足用户对关注业务方向实现动态感知与 QB 智能分析及实现业务的需求支撑。

1.2范围

此份文档的阅读对象包括：本项目的所有开发人员，测试人员，项目经理，以及 QA 小组人员。

2 总体概述

2.1系统概述

平台通过对行业数据和开源数据进行融合集成，可用于辅助业务部门对特定事件、特定主题、特定领域等方面数据进行数据监测与分析，筛查汇总散布于海量数据中的重点信息，结合知识库对重点信息进行图谱分析及关联挖掘，深究数据包含的价值信息，最终为业务部门的业务需求提供保障。

2.2用户群体

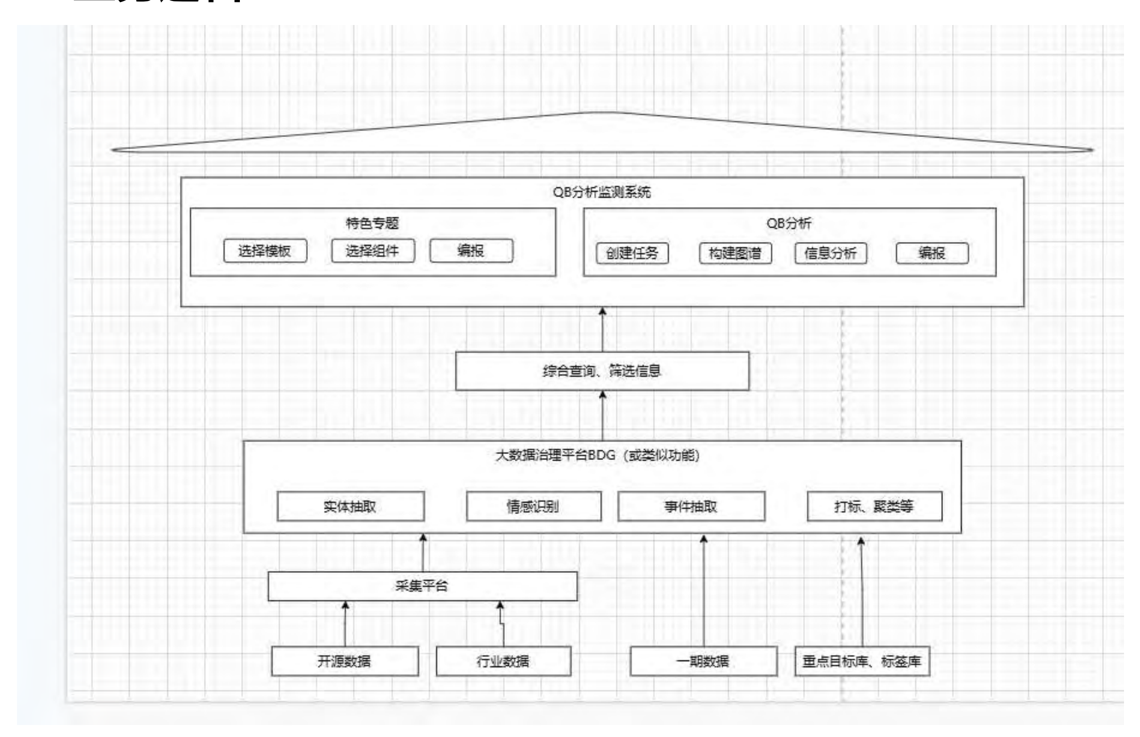
系统面向 QB 分析师和甲方业务部门人员，适用于初级用户和资深用户。

2.3功能清单

编号	一级模块	二级模块	功能点	子功能点
1	综合搜索	——	搜索快捷入口	支持多模态检索、支持图谱检索
2		——	搜索结果查看和处理	支持二次搜索、高级搜索，支持信息的收藏、导出、推送到 QB 分析，复制链接，复制信息等功能。
3		——	支持 S 网 A 网的结果查看	
4		——	搜索信息的分析	支持信息作者画像的查看，支持查看相关地点、人物、机构。支持外文翻译
5	特色专题	专题管理	专题列表	支持专题的分类管理，支持专题的增删改查。
6		——	专题信息列表	支持对专题筛选后的信息列表进行查看。
7		——	专题分析	支持对专题进行多维度的图表分析，支持生成分析报告。
8		模板管理	支持对专题模板的管理，包括系统模板、个人模板。支持对模板的增删改查。	
9	QB 服务	任务管理	支持对 QB 任务的增删改查，支持对人物配置人员、配置	

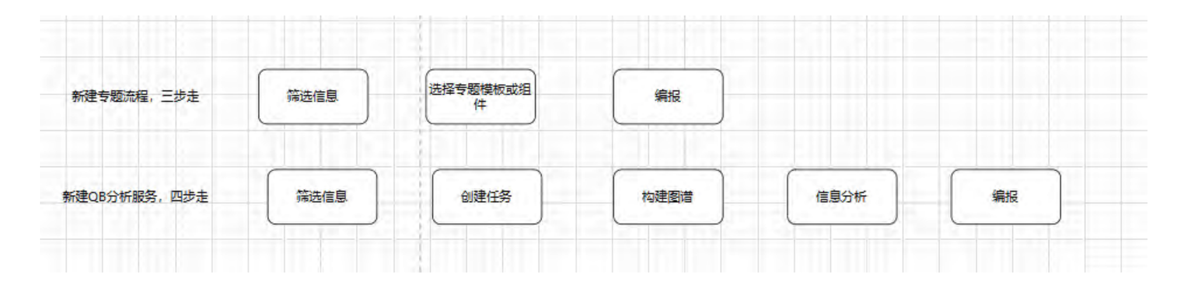
			分析结果，生成分析报告。	
10		图谱管理	支持对图谱的增删改查。	
11		信息分析	关联分析	支持对图谱数据进行关联分析，支持保存中间分析结果和截图。
12			时空分析	支持对图谱数据加载到地图上，进行时空分析。支持保存中间分析结果和截图。
13	个人中心	我的收藏	支持查看个人收藏的信息，支持导出、复制，推送到关联分析。	
14		我的结果	支持查看个人的分析结果，包括中间分析结果，截图。	
15		修改资料	复用公司统一总控平台	
16		修改密码	复用公司统一总控平台	
17		退出登录	复用公司统一总控平台	
18	报告管理	——	支持对报告的查询、搜索、预览、下载等功能。	
19	采集概览	——	支持对采集情况的宏观统计展现。	
20	系统管理	——	复用公司统一总控平台，支持机构管理、用户管理、角色管理。	

2.4业务逻辑



整个产品的业务逻辑是围绕采集信息，进行特殊专题监测和 **QB** 服务分析开展的。

2.5业务流程图



2.6功能结构图



2.7核心功能

系统核心功能模块包括综合搜索、特色专题、QB 服务。

（1）综合检索：支持用户对采集信息的多维度查询，以及对信息的深层分析。

（2）特色专题：基于采集数据，提供大量业务组件和模板，辅助初级用户快速搭建分析专题，生成分析报告。

（2）QB 服务：基于采集数据，提供多维度图谱分析算子，辅助资深用户去挖掘时间背后的人事组及关系。

2.8角色说明

角色	所属部门	权限说明
普通用户	待后续调研机构树	数据权限严格管控，只能看到个人数据+公共专题
管理员	待后续调研机构树	只能看到管理员个人数据，可修改公共专题

3 功能需求说明

统一说明：

功能需求中，对于业内通用功能将在此集中说明。对于特殊情况，以具体章节描述为准，否则全部采用统一说明的要求。

1、交互规范

1.1 页面尺寸

页面采用 1920*1080 大小设计，其他屏幕分辨率依照规则进行相应的尺寸适配。

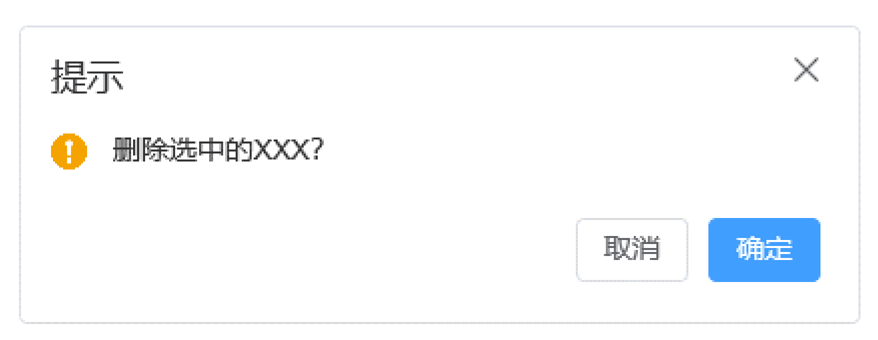
1.2 悬停效果

所有菜单导航栏、分类树和操作按钮，鼠标划过、悬停时均需支持变色提示。参考公司 OKS 产品：http://222.190.246.206:40037/oks/#/ontology/concept

2、操作规范

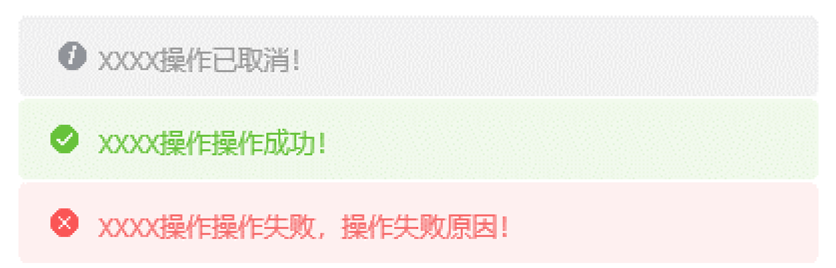
2.1 删除操作规范

所有删除操作均需弹窗进行二次确认。XXX 为系统中专题，任务，图谱等新增内容。



2.2 操作结束提示词

按照常用消息类型，将提示分为取消操作、操作成功、操作失败 3 类。

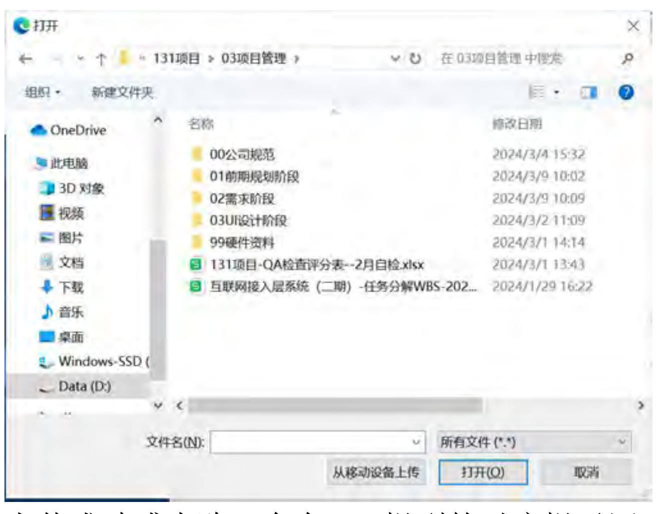


对于删除操作，删除成功后，会提示“删除操作成功。”

对于勾选后生效的按钮，若用户在未勾选下，直接点击按钮，则会提示“操作失败，未选中任何信息。”

2.3 上传文件交互规范

点击相关区域，弹窗选择文件，需进行文件格式、文件大小校验。校验内容见文件上传提示词。如下图所示。



上传成功或失败，会有 2.2 提到的对应提示词。

3、表格规范

3.1 表格排序

表格中无论是否包含创建时间，默认情况均按照创建时间降序排序，除非特殊说明；

编辑功能不改变表格中的数据展示顺序；

列表项内容包含数字、日期的，表头默认增加排序；



3.2 表格内容字数限制

列表项文字过长如何显示？当列表项的宽度无法完整显示所有文字时，后面加 3 个点，鼠标移入后显示全部。



3.3 分页器规范

共 78 条　10条/页　‹ 1 2 3 4 5 6 ••• 8 ›　前往 1 页

分页器于列表搭配使用，当数量超过 10 个时，列表底部显示分页器。支持每页 10 条、20 条、50 条、100 条切换。

4、查询搜索规范

4.1 输入字符限制

单行输入框输入字符限制为 20 个。

多行输入框输入字符限制为 200 个。

4.2 所有查询搜索框、下拉筛选框，均支持模糊搜索和一键清空。

3.1综合搜索

登录页面采用公司通用组件账号密码登录。

参考 OKS：http://222.190.246.206:40037/oks/#/login

登录后系统默认展示综合搜索内容，其包括搜索首页、搜索结果页和信息分析页。

3.1.1 搜索首页

➢ 原型如下：



➢ 页面布局：分为 Header 栏、Main 栏。

Header 栏为导航栏，从左至右依次为“Logo+系统名称”、“综合搜索”、“特色专题”、“QB 服务”、“报告管理”、“系统管理”；右侧为登录个人账号信息。

➢ Header 栏：

一级菜单	备注	二级菜单	备注	
综合搜索	单击，切换至综合搜索页面	无		
特色专题	单击，切换至特色专题首个二级菜单（专题管理）	专题管理、模板管理	单击，跳转至相关页面	
QB 服务	单击，切换至 QB 服务首个二级菜单（任务管理）	任务管理、图谱管理、QB 分析	单击，跳转至相关页面	
报告管理	单击，切换至报告管理页面	无		
系统管理	单击，切换至系统管理页面。需管理员角色权限才能查看。	用户管理、角色管理、单位管理、	单击，跳转至相关页面	
个人账号	单击，向下展开二级	我的收藏、我		

	菜单	的成果、修改资料、修改密码、退出登录		

➢ Main 栏：综合搜索入口。

按照最小可行性产品原则（MVP 原则），所有搜索方式，暂不支持历史搜索记录功能。

（1）文本搜索

点击【文本搜索】图标，页面如下图所示。

支持按照关键词搜索、高级搜索。提示词“请输入关键词搜索”，输入后点击搜索图标，进入【文本搜索结果页】。



点击【高级搜索】按钮，页面如下图所示。点击搜索，将筛选条件带入【文本搜索结果页】。点击返回，则隐藏高级搜索弹窗。



分类维度及取值如下：

维度	取值	备注
渠道	新闻媒体、社交平台、短视频、S 网 A 网、其他通道	单选。 新闻媒体：中国记协会员名录_百度百科(baidu.com) 国外新闻网站：BBC、CNN 社交平台：微信，微博，推特等 短视频：抖音，快手，YouTube S 网 A 网：特定网站；
情感	正面、中性、负面	单选。
地域	境外+中国 34 个省级行政区	单选。
事件大类	Gdlet20 个事件大类	单选。如下图

是否噪音	正常信息、疑似噪音	单选。噪音是指传播学中无用、虚假及有害的信息。详见公司噪音算法。
评论数	100 以下； 100--1000； 1000 以上	单选。

备注：
Gdlet20 个事件大类如下。

代码	含义	代码	含义	代码	含义
01	发表公开声明	08	投降	15	武力表明立场
02	上诉	09	调查	16	减少关系
03	明确表示合作	10	需求	17	胁迫
04	咨询	11	指控	18	袭击
05	从事外交合作	12	拒绝	19	战争
06	参与材料合作	13	威胁	20	使用非常规暴力
07	提供援助	14	抗议		

参考来源：Gdelt 数据库初探 (douban.com)

（2）图片搜索
点击【图片搜索】图标，通过图片来搜索视频信息。页面如下图所示。
支持选择图片或拖拽图片。图片格式及大小限制，见页面提示词。上传成功后，跳转到【图片搜索结果页】。



（3）跨模态搜索
点击【图跨模态搜索】图标，通过文本来搜索视频信息。页面如下图所示。
支持按照关键词搜索。提示词“请输入关键词搜索”，输入后点击搜索图标，进

入【跨模态搜索结果页】。



（4）视频搜索

点击【视频搜视频】图标，通过视频来搜索视频信息。页面如下图所示。

支持选择视频或拖拽视频。视频格式及大小限制，见页面提示词。上传成功后，跳转到【视频搜索结果页】。



3.1.2 搜索结果页

页面布局为上下结构，顶部是四种搜索方式，其交互与搜索入口页一致，区别在于文本搜索时，能够查看筛选条件。

3.1.2.1 文本搜索结果

文本搜索结果是对所有采集通道结果的信息展现。



➢ 对于多个信息，支持功能如下：

功能点	功能描述	前置条件	备注
勾选全选功能	支持对信息的勾选、全选。选中后，支持推送到 QB 分析、收藏、导出操作。	无	
去重功能	支持对信息进行去重或不去重展现。去重是指信息标题和作者完全一样。	无	单选下拉，默认显示为不去重
信息排序	支持信息按照如图下拉框条件进行排序。 默认显示为发布时间降序，保证最新事件信息在顶部。		发布时间降序 发布时间升序 转发数降序 转发数升序 评论数降序 评论数升序
推送到 QB 分析	将勾选的信息所涉及的节点和关系，一起推送到关联分析中。相当于批量新增节点和关系。	勾选信息	若未勾选直接点击按钮，则提示操作失败，未选中相关信息。
收藏	将勾选的信息进行收藏，在个人中心的"我的收藏"进行查看。	勾选信息	若勾选信息中包括已收藏和未收藏的信息，此时加入收藏时，在我的收藏中自动去重。
导出	将勾选的信息按照所需信息进行导出，支持导出格式为 Excel。 导出设置如图所示。	勾选信息	导出设置



➢ 对于单个信息，支持功能如下：

功能点	功能描述	备注
封面图	单一封面图显示	
情感倾向	对信息进行情感分类，包括正面、中性、负面。	

信息类型	包括原发、转发、评论。	
标题	显示信息标题。命中关键词的部分，字体标红突出显示。	点击标题进行信息分析页。
是否已读	用户点击后，会在标题后面打标显示已读。	
信息发布时间	格式为年月日时分秒	是采集时作者发布的当地时间
正文	显示信息正文。页面显示首个命中关键词的部分，字体标红突出显示。前后超出字数则用…显示。	悬停时无效果。
信源类型	包括网站、电子报、新闻客户端、电视媒体、贴吧、论坛、视频、社交平台、其他	所有类型，包括境内外信源通道。
信源名称	格式为：通道名称+作者信息（或板块信息）。 （1）搜狐网—军事 （2）抖音—董宇辉 （3）北京日报	若无作者信息，则不显示，如示例 3。
赞转评信息	显示信息的点赞、转发、评论的数量信息。	
查看原文	点击后，在新的浏览器页签查看原文	五个图标，对应五个功能。鼠标悬停在图标时，显示文字。
加入收藏	点击后，该条信息进入我的收藏。按钮改为“取消收藏”。	
素	点击后弹窗导出设置，同批量导出。	



材导出		
复制链接	复制原文链接信息	
复制信息	复制该信息在列表展现的元数据。示例如下 标题：转发微博 链接： http://weibo.com/5671712418/O4qVV6TeK?refer_flag=1001030103_ 摘要：空 时间：2024-03-11 02:41:18 来源：微博 作者：PeterDong 的微博 倾向性：中性	

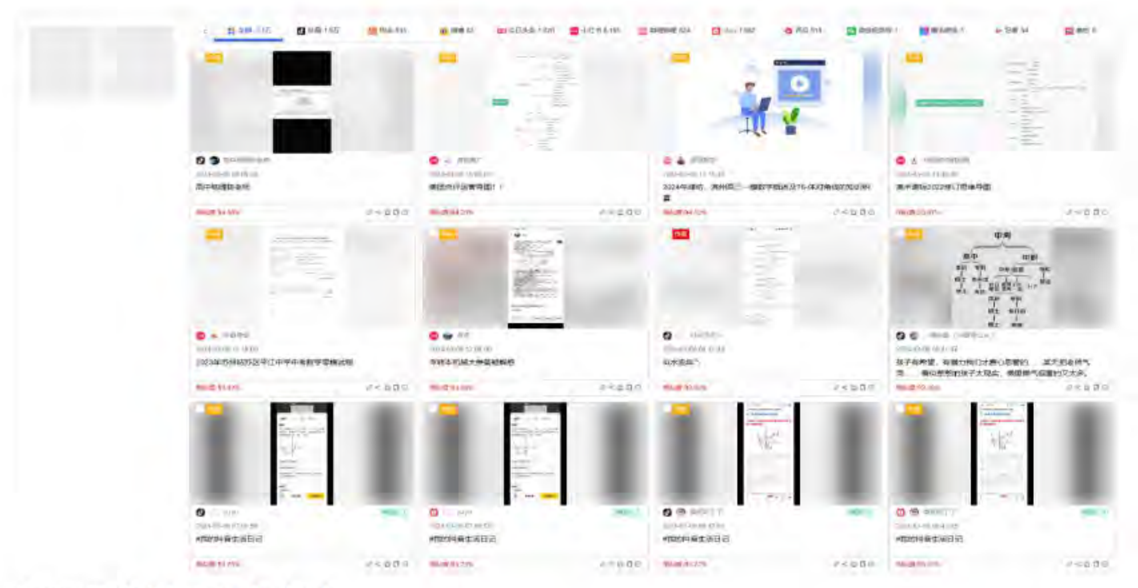
➢ 特殊情况说明



（1）对于外文信息，增加【翻译】按钮。点击【翻译】后，将标题和正文翻译为中文展示。
（2）对于 S 网 A 网信息，由于其是电商网站结构，会增加价格信息，去掉赞评转信息。其余不变。

3.1.2.2 图片搜索结果

图片搜索结果是对视频信源采集通道结果的信息展现，简单理解为图搜视频。



每行最多展示 4 个卡片。

- 对于多个信息的操作，与文本搜索功能一致。
- 对于单个信息，视频信息功能如下：删除内容，标黄内容是方便与文本搜索结果进行对比。

功能点	功能描述	备注
封面图	单一封面图显示	
情感倾向	对信息进行情感分类，包括正面、中性、负面。	
~~信息类型~~	~~包括原发、转发、评论。~~	
标题	显示信息标题。命中关键词的部分，字体标红突出显示。	点击标题进行信息分析页。
~~是否已读~~	~~用户点击后，会在标题后面打标显示已读。~~	
信息发布	格式为年月日时分秒	是作者发布的当地时间

时间		
正文	~~显示信息正文。页面显示首个命中关键词的部分，字体标红突出显示。前后超出字数则用...显示。~~	悬停时无效果。
信源类型	包括网站、电子报、新闻客户端、电视媒体、贴吧、论坛、视频、社交平台、其他	所有类型，包括境内外信源通道。
信源名称	格式为：通道名称+作者信息（或板块信息）。 （1）搜狐网—军事 （2）抖音—董宇辉 （3）北京日报	若无作者信息，则不显示，如示例 3。
相似度	利用图片相似度算法，显示视频中每帧图片相似度的最高值。	
~~赞转评信息~~	~~显示信息的点赞、转发、评论的数量信息。~~	列表页不展现，分析页展现
查看原文	点击后，在新的浏览器页签查看原文	五个图标，对应五个功能。鼠标悬停在图标时，显示文字。
加入收藏	点击后，该条信息进入我的收藏。按钮改为“取消收藏”。	
素材导出	点击后弹窗导出设置，同批量导出。	
复制链接	复制原文链接信息	
复制信息	复制该信息在列表展现的元数据。示例如下 标题：XXX 链接： http://douyin.com/5671712418/O4qVV6TeK?refer_flag=1001030103_	

	摘要：空 时间：2024-03-11 02:41:18 来源：抖音 作者：董宇辉 倾向性：中性	

3.1.2.3 跨模态搜索结果

与图片搜索类似，可以理解为文本搜视频。差异点在于输入内容改为文本，相似度算法改为文本-视频相似度算法。

3.1.2.4 视频搜索结果

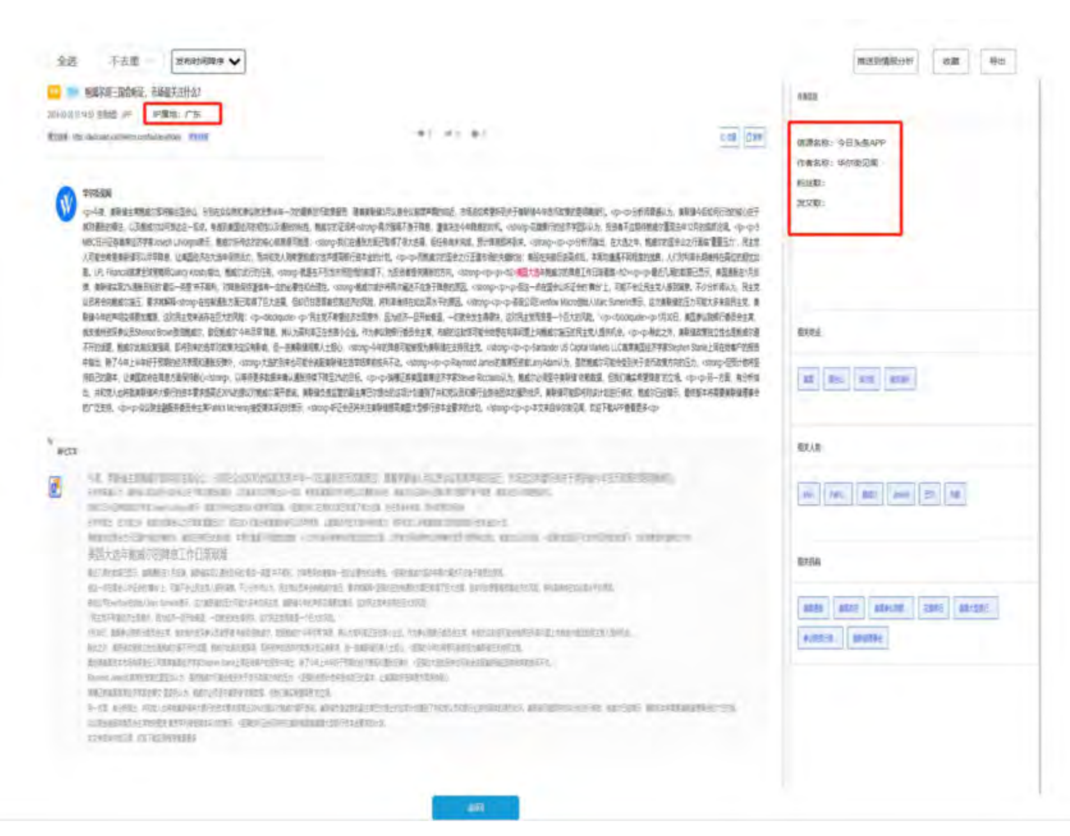
与图片搜索类型，可以理解为视频搜视频。只是输入内容改为视频，相似度算法改为视频相似度算法。

3.1.3 信息分析页

信息分析页主要包括 3 大类。图文信息分析、视频信息分析、S 网 A 网信息分析。

3.1.3.1 图文信息分析

对于图文信息分析，其分析维度如下图所示。



除去列表页中的元数据信息，额外新增了 IP 属性，右侧的作者信息中，新增粉丝数，发文数。新增相关地点、相关人物、相关机构信息。

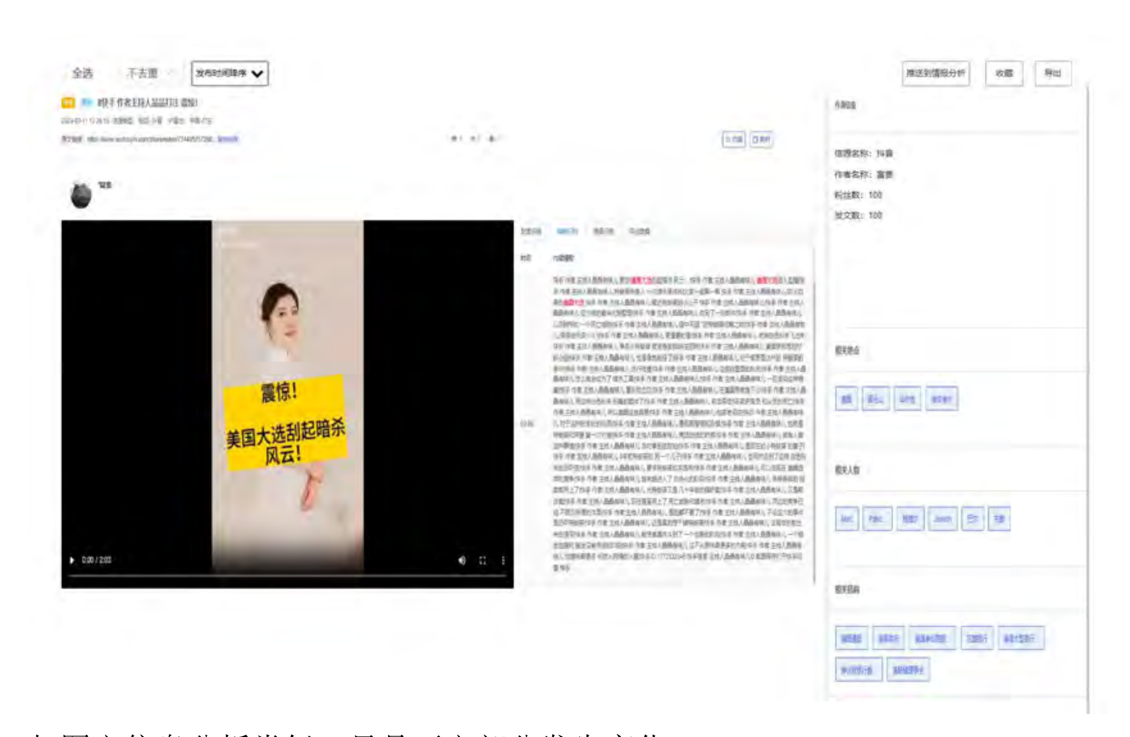
点击翻译，则在下文显示对应的中文信息。对于中文原文，则弹出提示词“原文

信息已经是中文”。

若某篇信息无法抽取相关内容，则只显示表结构信息，没有数据信息。
点击返回按钮，返回到列表页。

3.1.3.2 视频信息分析

对于视频信息分析，其分析维度如下图所示。

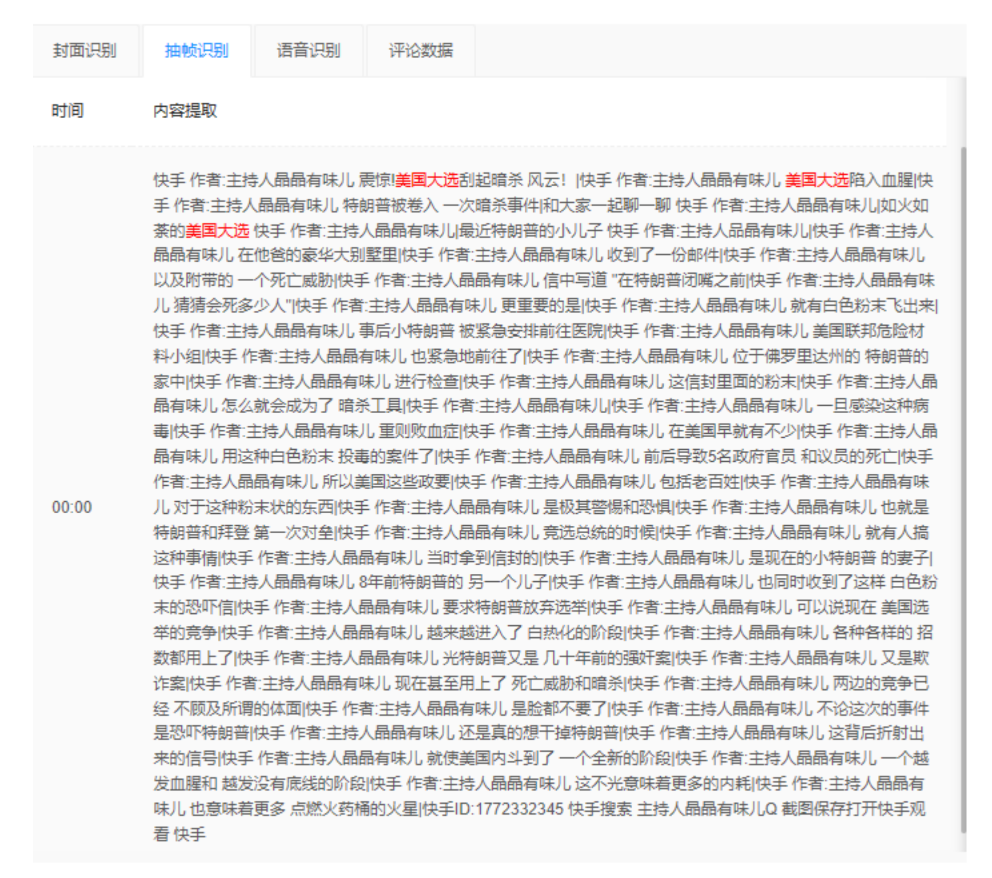


与图文信息分析类似，只是正文部分发生变化。
正文包括左侧的视频和右侧的识别。视频支持播放，控制音量、全屏和下载。右侧的是不包括封面识别、抽帧识别、语音识别和评论数据。

封面识别：
利用 OCR 技术，将封面中内容提取出来，字体标红突出显示涉及监测词的内容，并显示时间。



抽帧识别
利用 OCR 技术，将不同帧中的内容提取出来，字体标红突出显示涉及监测词的内容，并显示时间。



语音识别：

利用语音识别技术，将语音中的内容提取出来，字体标红突出显示涉及监测词的内容，并显示时间。



3.1.3.3 S 网 A 网信息分析

对于 S 网 A 网的信息分析，其分析维度如下图所示。



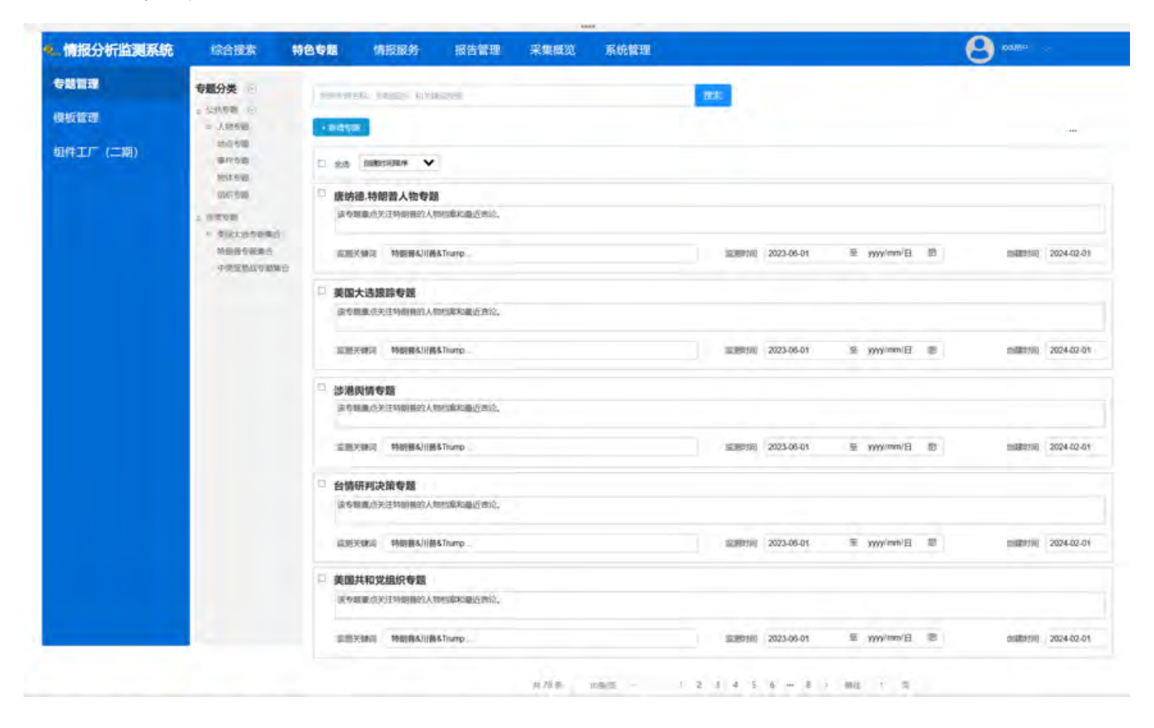
除去列表页展现的元数据信息，额外新增了商品类型，数据来源、可支付币种，信源地址，详情、评价，和留言等信息。
备注：不同的特殊网站，表结构信息可能不同。缺少的内容展示为空。本文是以公司现有能力举例。

3.2特色专题

包括专题管理和模板管理。左侧导航栏可以切换专题管理和模板管理页面。

3.2.1 专题管理

➢ 原型如下：

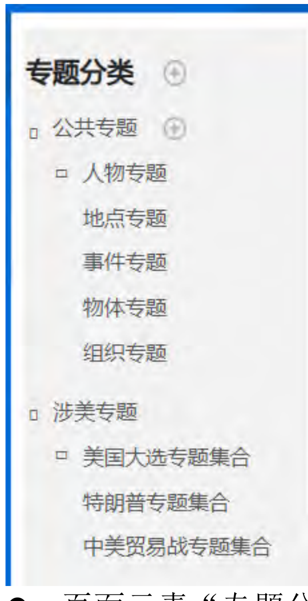


➢ 入口：点击左侧导航栏“专题管理”。左侧导航栏支持展开、收起功能。

3.2.1.1 专题列表页

➢ 主要功能区分为左右结构。

左侧是专题分类树，是纯粹的分组，不包括专题名称。



- 页面元素“专题分类”，默认为根节点。点击右侧加号，新增子节点。分类数量有限，小于 1000，暂不考虑树节点过多造成的加载时间问题。
- 页面元素“公共专题”，默认为首个二级节点。只有管理员角色具有公共专题的编辑权限。普通用户只能查看公共专题内容。专题来源，一方面是管理员新增，另一方面是普通用户可以把自己专题发布共享。
- 页面元素“涉美专题”，是当前用户个人新增的二级节点。点击节点，显示该节点下所有的专题。点击父节点，显示所有子节点的专题。点击专题分类时，显示该用户能查看的全部专题。
- 专题属于敏感信息，需要数据权限管理。每个用户只能查看个人专题和公共专题，只能编辑个人专题。
- 用户点击新增专题，若没有选择左侧分类树，则默认为属于根节点专题分类。

➢ 右侧是通用的卡片列表形式，暂不支持切换为卡片平铺形式。



上方是搜索框，支持按照专题名称、专题描述、关键词进行搜索。

中间是列表的功能操作区，支持专题列表的增删改查、全选、排序。

当选择多个，点击重命名、修改时，会有提示词“操作失败，只能选择单个”

➢ 下方是卡片列表信息。

功能点	功能描述	备注

专题名称		单行输入框字符限制，按统一规范执行。
专题内容		
监测关键词		
监测时间		格式为日期
创建时间		格式为日期

3.2.1.2 新增专题页

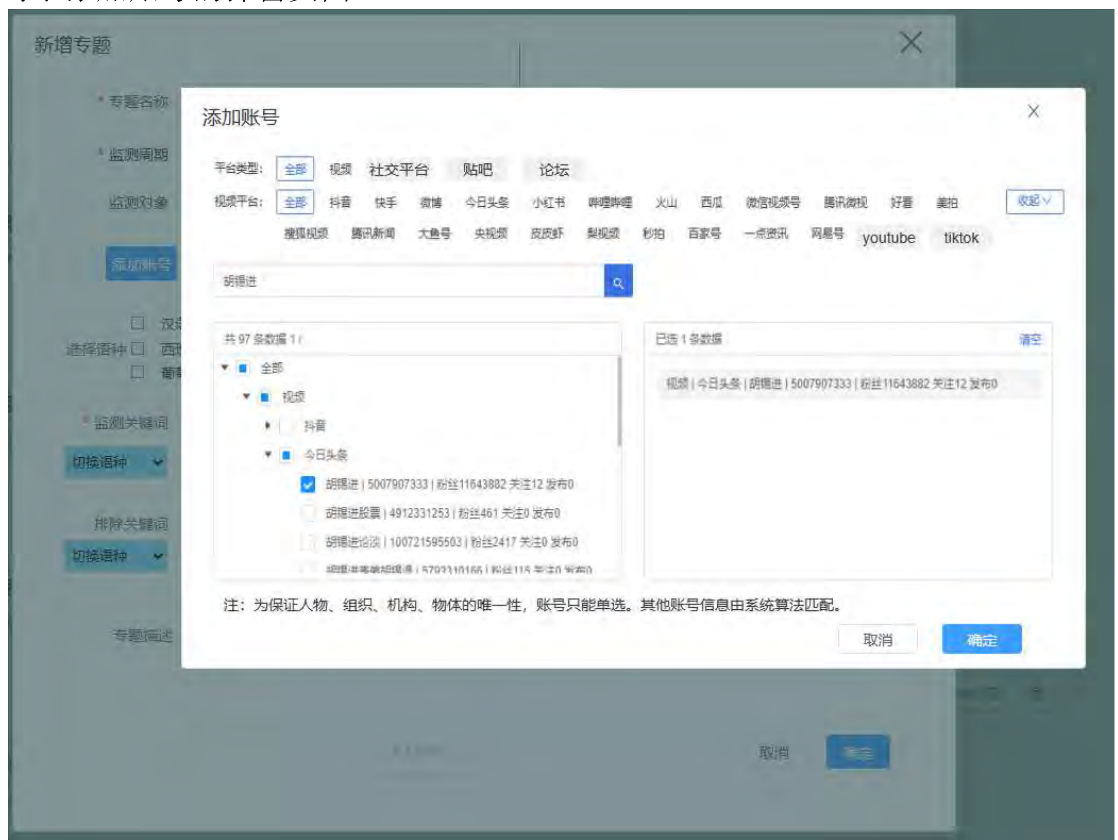
新增专题的原型如下：



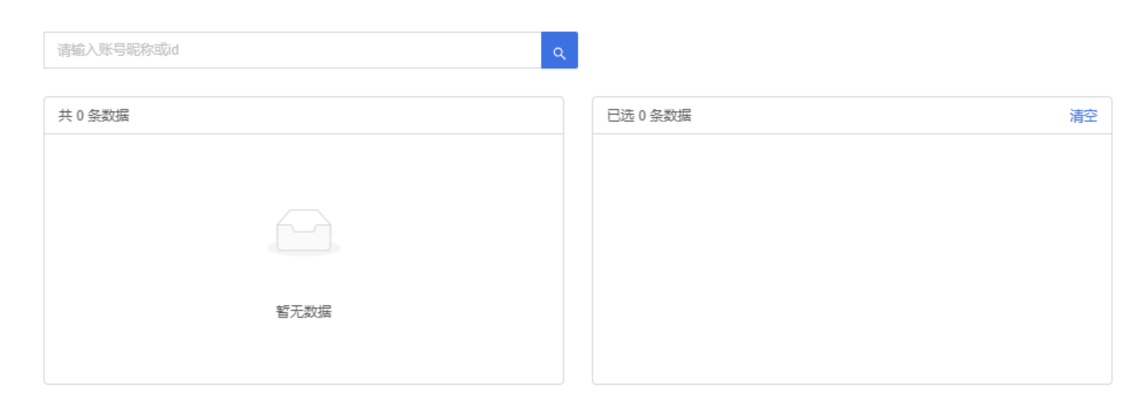
字段名称	默认值	字段类型	是否开放修改权限	是否是必填项	备注
专题名称		文本	是	是	按照统一规范要求：不超过 20 个汉字
监测周期		日期	是	是	
监测主体	显示提示“为了监测效果，请确保唯一性”	文本	是	是	❓帮助提示词“监测主体请确保没有歧义，与监测组件逻辑相关”

添加账号	--	--	是	否	事件类专题，无须账号。 点击按钮后弹窗。
选择语种	默认勾选中文	--	是	否	可多选 12 大语种。 勾选多语种后，将把全部的译文作为监测词去筛选采集信息。
监测关键词	--	--	是	是	支持下拉选择标题+正文、标题、正文； 支持与或优先操作； 支持切换语种；
排除关键词	--	--	是	否	同上
专题描述	无	文本	是	否	
关键词使用示例	--	--	--	--	前端写死，辅助用户写提示词。 后续根据业务调研补充。

➢ 对于添加账号的弹窗页面。



- 账号平台类型是采集信源类型的子集，如图所示，不包括网站，电子报等。都包括境内外信息。
- 每一类平台的具体采集信源平台，后续与业务调研后持续补充。
- 搜索框初始状态，见下图



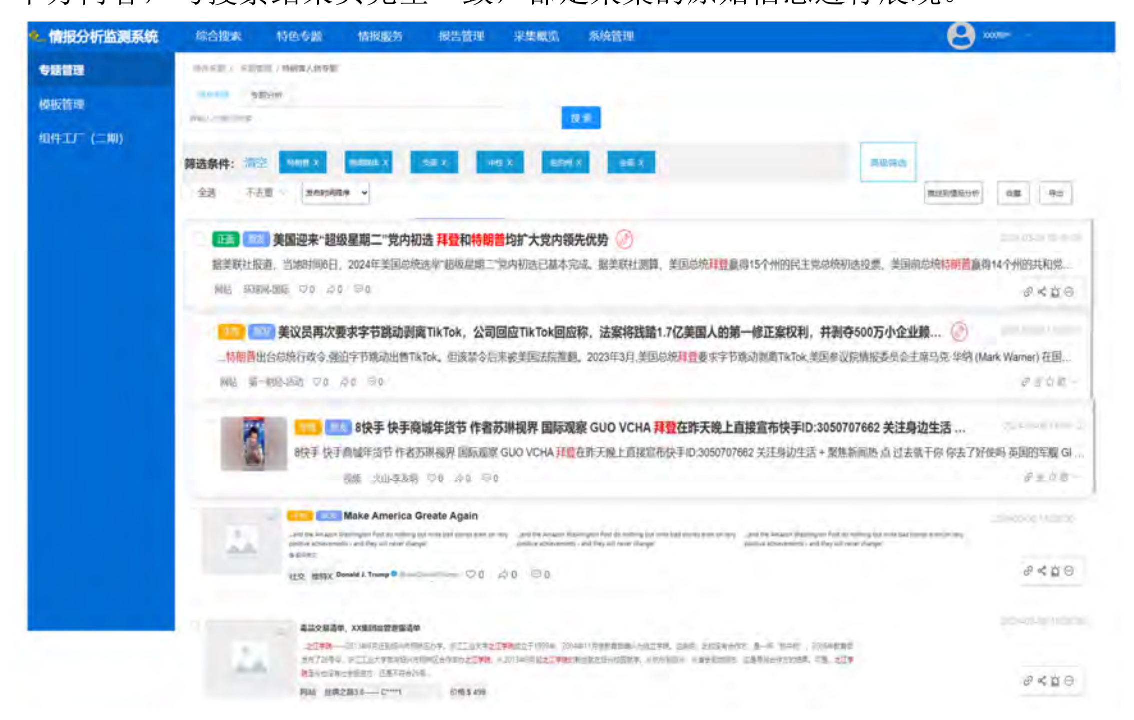
- 输入账号后 ID 后，显示如图所示。将多选框改为单选框。

3.2.1.3 专题详情页--信息

在列表页，点击专题名称，进入专题详情页的信息列表页，方便用户通过采集结果来验证监测词是否有效。

顶部是面包屑导航。两个页面“信息列表”“专题分析”可点击切换。

下方内容，与搜索结果页完全一致，都是采集的原始信息进行展现。



3.2.1.4 专题详情页--分析

专题分析页面包括左右两部分。



➢ 左侧是组件工厂

组件工厂，是系统内置的业务组件。本期不支持组件的二次开发。组件分类固定为下拉框展示内容。当选择分类后，组件内容随之改变。

当用户拖拽组件时，自动将组件所需数据显示出来，无须用户去配置组件的数据来源。技术支撑是系统内置采集后的人物库、机构库和各种抽取结果。

图形类组件，拖入后自带导入、导出和删除图标。

文本类组件，拖入后自带编辑、导出和删除图标。



➢ 右侧是专题布局内容

页面初次加载时，是白板形式。

中间有提示词“暂无组件，请拖拽左侧组件进行添加，或者选择模板”。

下方有模板推荐，模板卡片样式为封面图+模板名称。点击某个模板，相当于批量拖入一堆组件。点击“更多模板”卡片，弹窗显示所有模板。



所有的专题，右上角都有操作区。支持将专题内容生成报告、存为模板和发布共享。

点击生成报告，需用户填写报告名称。



点击存为模板，需用户填写模板名称、模板描述。



点击发布共享，是将个人专题同步拷贝一份到公共专题，需用户选择分组。



3.2.1.5 人物类模板组件

选择组件后，将同步读取组件的分类信息，在右侧的目录树中展示。方便用户点击目录树内容，直接跳转过去。

备注：所有的组件，都是针对监测主体来设计。



（1）人物概述组件



- 该组件显示人物的头像、人物名称和人物简介。
- 支持编辑文本，导出功能和删除功能。其中，所有组件的导出按钮，支持导出数据和导出图片。点击导出图标后，再次选择导出内容，
- 数据来源是采集平台积累的人物库。通过创建专题时的监测主体和账号，保障人物库中的唯一性。

对采集平台的要求：需持续抽取、更新 wikidata 数据，形成人物基础库、人物动态信息库，作为行业积累。

（2）人物基本信息组件



- 该组件显示人物的基本信息，如图所示。
- 支持编辑文本，导入、导出和删除。导出数据后，在 Excel 中可以修改数据或表结构。再导入专题。系统将所有表结构字段按顺序分成两组显示。
- 数据来源是采集平台积累的人物基础数据库。通过创建专题时的监测主体和账号，保障人物库中的唯一性。相同功能，不再重复描述。

（3）教育经历



（4）职业经历

（5）关联账号

备注：不同的平台，其 logo 不同。系统内置全部平台的 logo。但导出的 Excel 中，图标是用平台地址来替换。

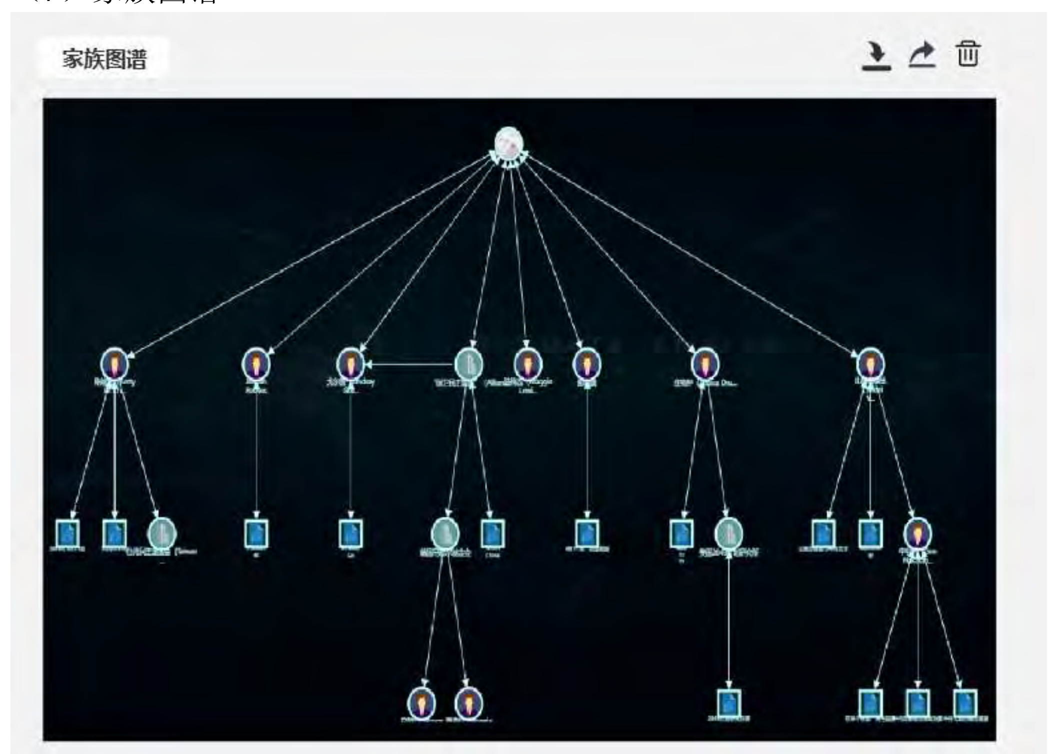
信息平台地址	账号名称	账号 ID	个人主页
www.youtube.com	Trump	798887388568	https:/youtube.com/Trump

（6）关联网站



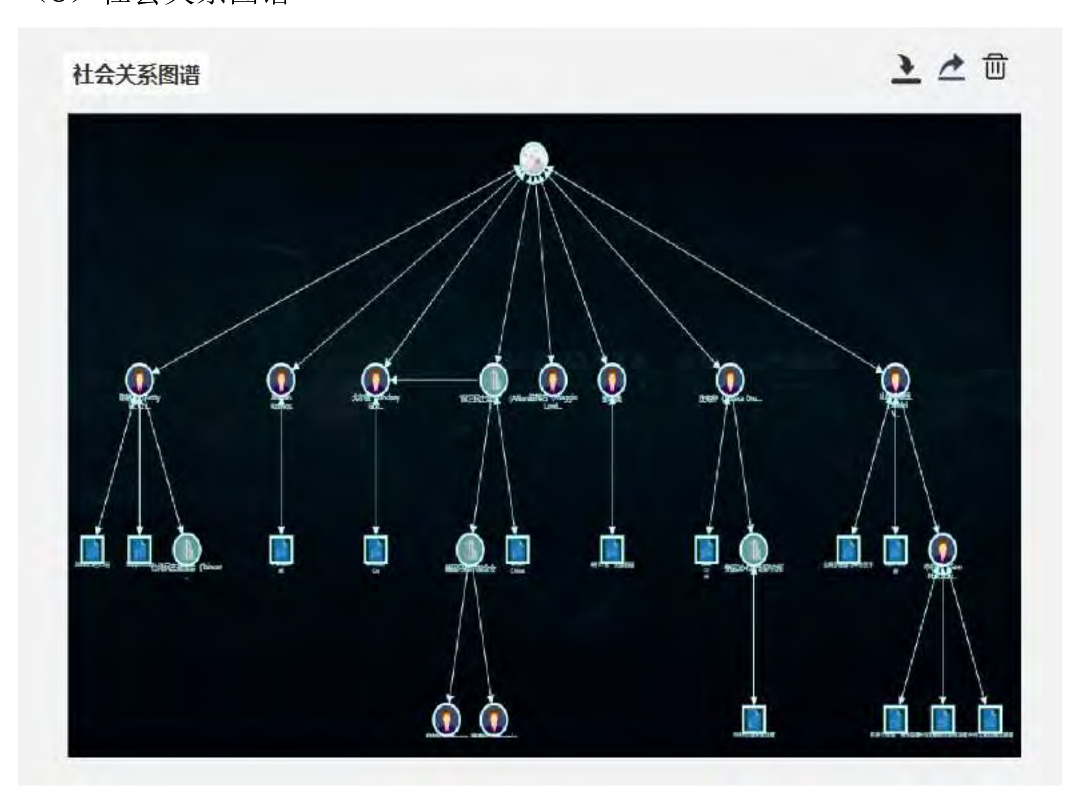
备注：与上图类似，但不同的网站，其 logo 保持一致。

（7）家族图谱



- 鼠标悬停在节点或关系时，显示节点名称或关系名称。
- 备注：导出图谱的模板，等后续技术反馈。
- 数据来源：人物基础数据库。

（8）社会关系图谱



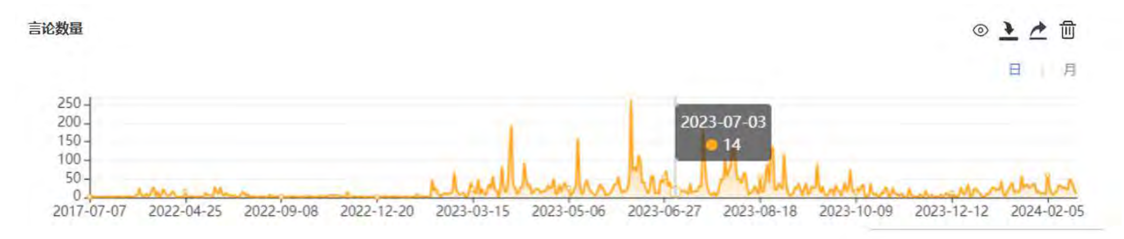
数据来源：专题筛选后的采集信息，经过社会关系抽取算法后形成人物社会关系

库。

待定信息：是算法提前处理，还是拖入组件时再处理。

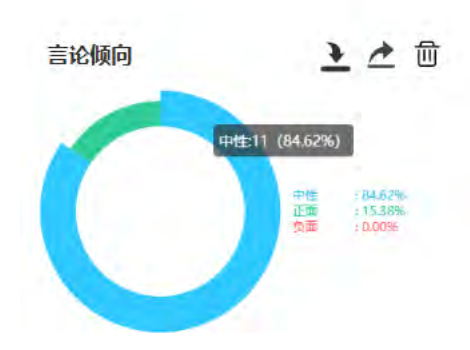
个人倾向是，产品提前梳理好重点人物，作为人物库的积累。
对于人物库里的专题，算法提前处理后，组件直接读取结果。
对于人物库之外的人物，则拖入组件时，实时调用算法抽取。页面提示处理中。

（9）言论数量



鼠标悬停时，显示日期和言论数量。支持按日、月切换。时间轴的起点，最早是监测周期的起点。
数据来源：专题筛选后的采集信息，经过言论抽取算法处理，再进行统计展现。

（10）言论倾向



数据来源：专题筛选后的采集信息，经过言论倾向识别算法处理，再进行统计展现。

（11）言论详情



- 支持关键词搜索，按言论时间降序排列。
- 显示日期和言论内容。
- 每屏显示 10 条（暂定，以 UI 设计条数为准），超出后上下滚动显示。

数据来源：专题筛选后的采集信息，经过言论抽取算法处理，再进行信息展现。

（12）话题

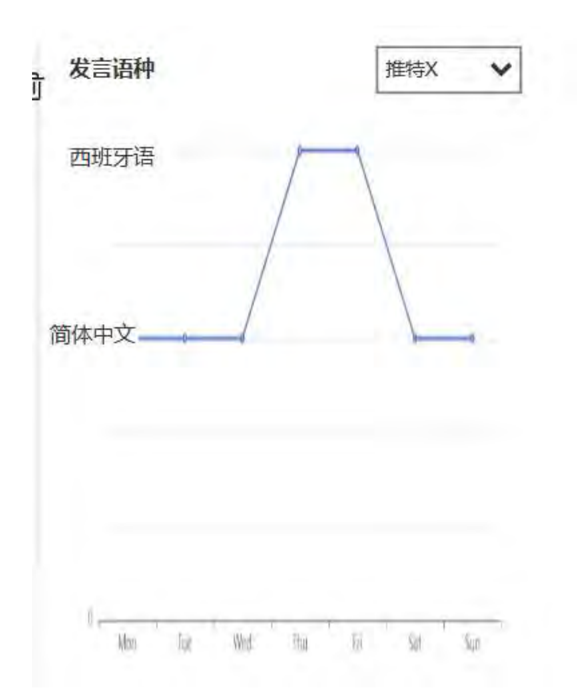
▾ 话题（30）

话题	数量
格鲁吉亚	127
乌克兰	80
俄罗斯	77
美国总统	67
支持者	66
德桑蒂斯	62
听证会	56
以色列	56
政治迫害	55
Mar_Lago	52
格鲁吉亚选举变态...	44
拜登政府	41

- 每屏显示 TOP10（暂定，以 UI 设计条数为准）。

数据来源：专题筛选后的采集信息，经过言论话题（主题）算法处理，再进行信息统计展现。

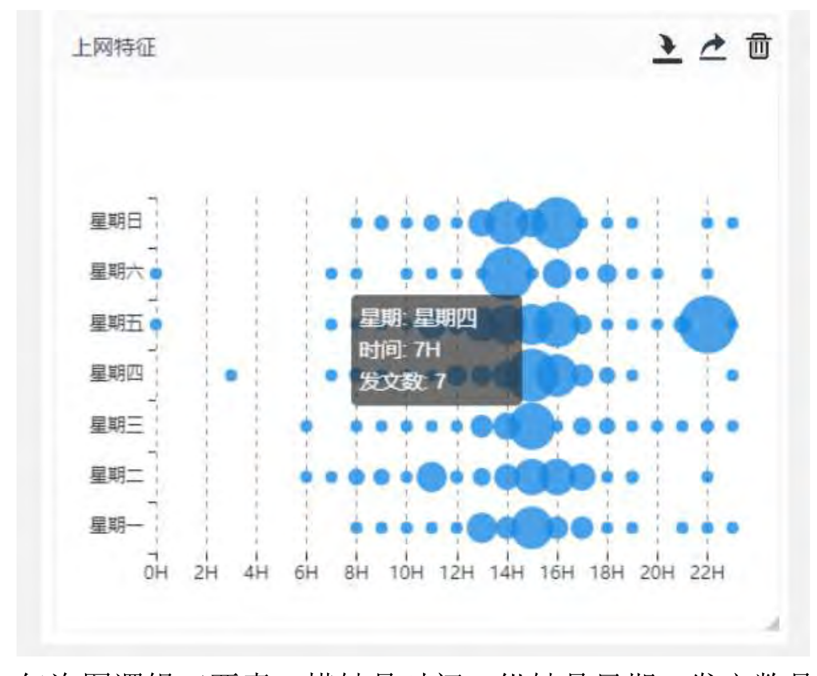
（13）发言语种



- 支持切换推特，Facebook，微博，微信，共计四种主流平台
- 横轴是日期排序，显示最近一月的发言语种。语种按照 12 大语种分类即可。

数据来源：专题筛选后的采集信息，经过言论语种算法处理，再进行信息展现。

（14）上网特征



气泡图逻辑三要素。横轴是时间，纵轴是日期，发文数是气泡大小。

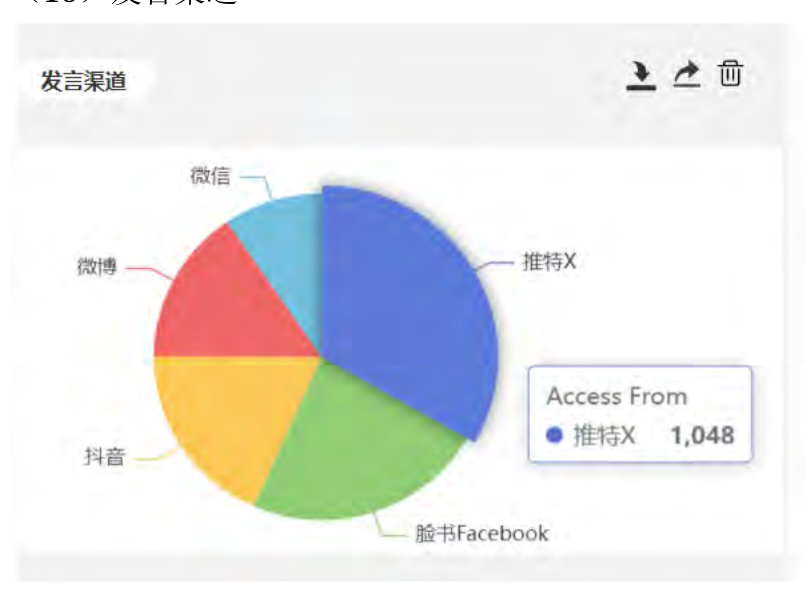
数据来源：专题筛选后的采集信息，经过言论时间算法处理，再进行信息统计展现。

（15）提及热词



数据来源：专题筛选后的采集信息，经过言论热词算法处理，再进行信息统计展现。

（16）发言渠道



数据来源：专题筛选后的采集信息，经过言论发言渠道算法处理，再进行信息统计展现。

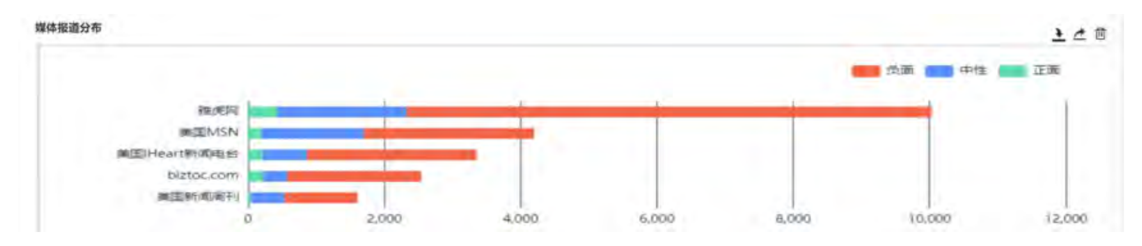
（17）相关活动



相关活动是复合组件，按照左 2 右 1 的布局。展示主体人物活动的时间发布，活动详情和活动类型。

数据来源：专题筛选后的采集信息，经过活动算法处理，再进行信息统计展现。活动算法相当于谓语提取。

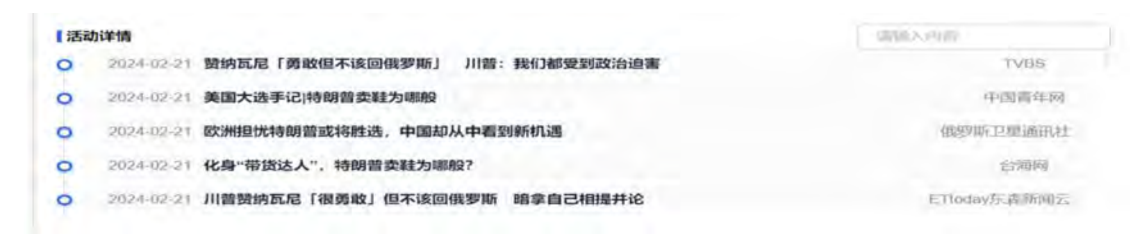
（18）媒体报道分布



- 每屏显示 TOP5（暂定，以 UI 设计条数为准）。

数据来源：专题筛选后的采集信息，经过来源媒体抽取处理，再进行信息统计展现。

（19）媒体报道详情



- 支持关键词搜索，按报道时间降序排列。
- 显示日期、报道内容、报道机构。
- 每屏显示 10 条（暂定，以 UI 设计条数为准），超出后上下滚动显示。

数据来源：专题筛选后的采集信息，经过来源媒体抽取处理，再进行信息展现。



在拖入第一个组件或者选择模板后，在底部会出现添加设置。可设置每行组件个数。最多是 3 个。

3.2.1.6 美国大选模板组件

整体大选模板的目录架构，如图所示

美国大选专题

- 大选热度趋势
  - （1）选举概述
  - （2）热点新闻
  - （3）选举日程
  - （4）事件热度
- 关注大选的人
  - （1）官方人物表态
  - （2）大V表态
  - （3）民众表态
- 关注大选的组织
  - （1）政府组织表态
  - （2）NGO表态
  - （3）组织言论热词
- 选举流程
  - 党内初选
  - 全国党代会
  - 总统竞选
  - 全民投票
  - 总统就职

1、选举热度趋势

包括 4 个组件。

（1）选举概述

选举概述

2024年美国大选”是美国举行的总统选举活动，美国实行总统制，每四年举行一次总统选举，总统任期4年。美国2024年总统选举时间为2024年11月5日

提醒：文本类组件，支持编辑，导出、删除。重复功能不再描述。

公司没有事件库，随着项目建设，开始积累 AQ 领域的事件库。

数据来源：
组件标题为“选举概述”，
文本框内容固定为以下格式。

组件结构	数据来源	示例	备注
年份	创建时系统时间	2024 年	
监测主体	创建专题时必填	美国大选	
是	组件写死		
国家/地区	对监测主体进行分词提取	美国	若分词库不好实现，则在创建专题时，增加地点信息。
举行的选举活动	组件写死		
美国实行总统制，每四年举行一次。美国 XX 年总统选举时间为	组件写死		由地点来触发组件词库，内置主要国家、地区的选举基本信息。比如韩国总统制，日本首相制。
2024 年 11 月 5 日	选举日程抽取		

（2）热点新闻



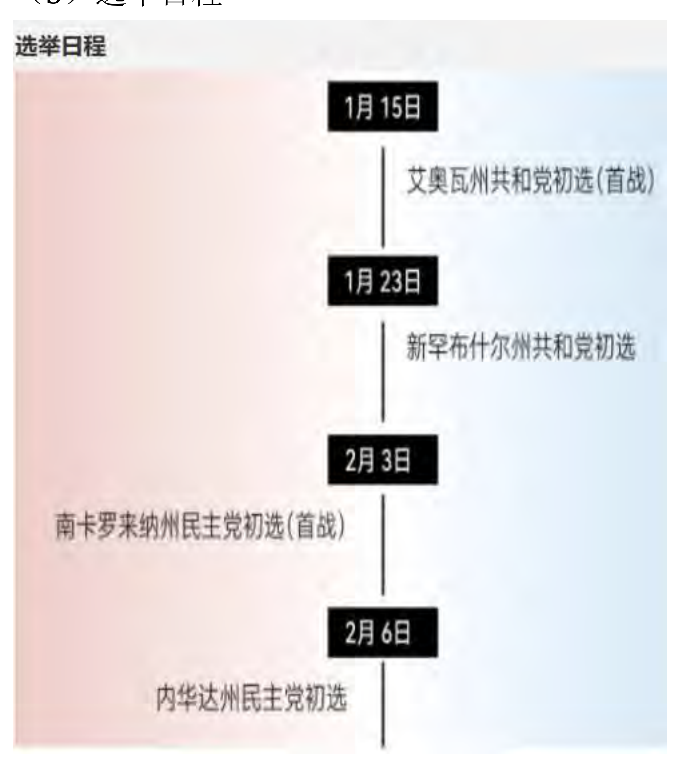
组件结构	数据来源	示例	备注
标题	写死	热点新闻TOP10	后续不再重复描述标题，见原型。
参选党派组件	单选下拉，内置国	民主党、共和党	内置主要国家的参选

	家与党派关系库。		党派。 二期再支持来源数据配置。
热点话题	多选下拉。20 个国家安全观+两派竞选纲领议题。	政治安全、国土安全、军事安全、经济安全、文化安全、社会安全、科技安全、网络安全、生态安全、资源安全、核安全、海外利益安全、生物安全、太空安全、极地安全和深海安全	下拉数据来源两部分。 二期再支持来源数据配置。
信息列表			点击弹窗，进入信息分析页。

数据来源：在专题监测词筛选下，根据采集时信息阅读量排序。默认显示 TOP10，可上下滑动展示。

备注：所有下拉组件，均支持多选操作，方便用户对比！组件选择后，本期不支持用户更改组件样式，比如不这次从折线图改为柱状图等。

（3）选举日程

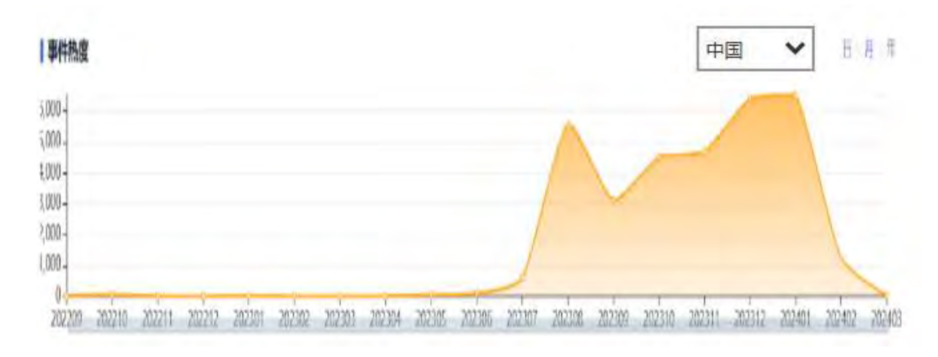


整体支持滚动显示，中间轴是日期，左侧是民主党，背景改为蓝色（蓝党）。右

侧是共和党，背景改为红色（红党）。

数据来源：要求采集平台，定向采集。积累到 AQ 行业事件库。
美国 ABC 新闻网站大选板块。
[选举新闻、直播视频和更新 - ABC 新闻 (go.com)]

（4）事件热度



事件热度：将专题筛选后的信息列表，按照时间进行统计展现。
纵轴是热度（数量），横轴是时间。支持切换日、月，没有年。

支持下拉选择国家。国家为五常+日韩。
再次备注：所有下拉组件，均支持多选操作，方便用户对比！

2、关注选举的人
（1）关注选举的政要排行

关注选举的政要排行

序号	人物	事件数	关注指数	
1	普京	251	3.53	言论 热词
2	欧盟	131	2.17	言论 热词
3	普京	110	2.15	言论 热词
4	岸田文雄	79	1.81	言论 热词
5	联合国	103	1.58	言论 热词

组件结构	功能	备注
序号		只显示 top5
人物	人物+头像	头像来自内置的人物库

事件数	统计周期内，标题+正文中包含人物（多语种）的信息列表数。	
关注指数	（个人事件数/前 5 事件总数）*10	占比情况，量程为 0-10，最多取两位小数。
言论热词	统计周期内，人物发言热词词云展示。	点击按钮，弹窗
统计周期切换	按日，按周，按月	去掉按年。

言论热词弹窗如下



（2）关注选举的政要排行

关注选举的时政大V分析

脸书 推特X 微博 微信

排行	相关大V	发文数	点赞	评论	转发	
1	胡锡进	100	100	100	100	粉丝画像
2	司马南	80	80	80	80	粉丝画像
3	卢克文	60	60	60	60	粉丝画像
1	胡锡进	100	100	100	100	粉丝画像
2	司马南	80	80	80	80	粉丝画像
3	卢克文	60	60	60	60	粉丝画像

组件结构	功能	备注
序号		只显示 top5
相关大 V	人物+头像	头像来自内置的人物库在监测词筛选下，将信息列表中的大 V 抽取出来。
发文数	显示总数	
点赞数	显示总数	
评论数	显示总数	
转发数	显示总数	
粉丝画像		点击按钮，弹窗。画像维度见下文。

统计渠道切换		境内外主流 4 个社交平台

粉丝画像弹窗如下：
画像维度包括性别，年龄段（10 年为一段），学历，职业、活跃时间段（按小时统计），单身比例。



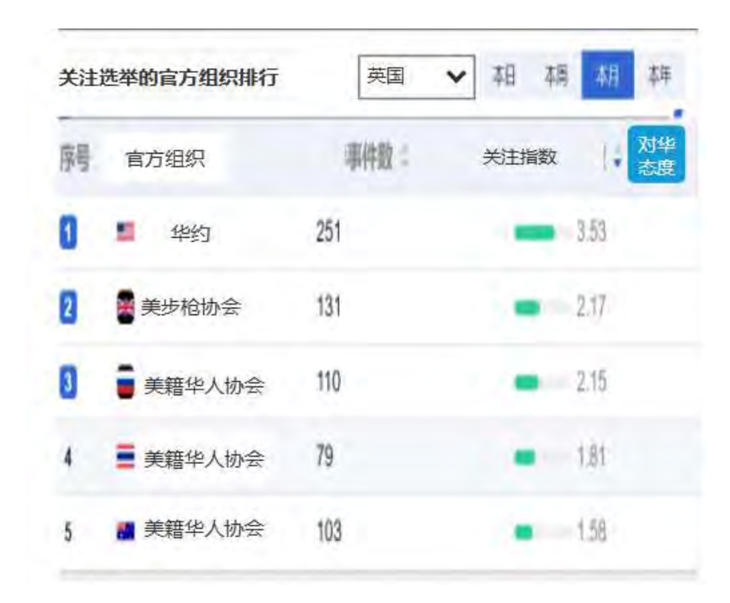
（3）社媒情感分析



在监测词筛选下，将信息列表中的信息情感抽取出来，并按四大主流渠道进行划分。按日统计展现。

3、关注选举的组织

（1）关注选举的官方组织排行



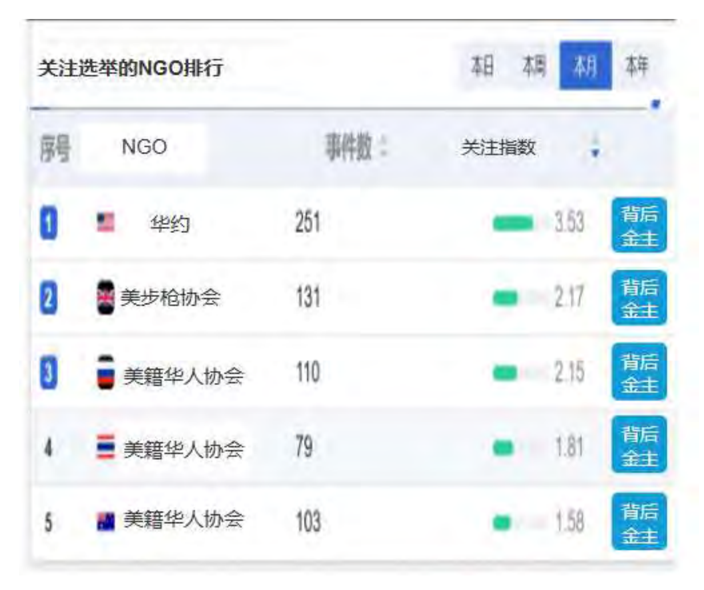
关注选举的官方组织排行

英国

序号	官方组织	事件数	关注指数	对华态度
1	华约	251	3.53	
2	美步枪协会	131	2.17	
3	美籍华人协会	110	2.15	
4	美籍华人协会	79	1.81	
5	美籍华人协会	103	1.58	

展现逻辑同上。
官方组织定义：公司已有组织库，比如美国步枪协会。后续与客户进行业务梳理，是否有重点关注组织。

（2）关注选举的 NGO 排行



关注选举的NGO排行

序号	NGO	事件数	关注指数	
1	华约	251	3.53	背后金主
2	美步枪协会	131	2.17	背后金主
3	美籍华人协会	110	2.15	背后金主
4	美籍华人协会	79	1.81	背后金主
5	美籍华人协会	103	1.58	背后金主

展现逻辑同上。
NGO 组织定义：公司已有组织库，比如美国国家民主基金会。后续与客户进行业务梳理，是否有重点关注组织。

（3）组织热词
政府组织和 NGO，发言中提到的词云。



4、党内初选

（1）初选投票统计



组件说明

- 支持两党切换展示
- 支持显示更新时间，美东。

- 支持实时显示候选人的投票数，降序展示。
- 支持显示提名线，超出提名线的候选人将成为党内候选人。

数据来源：对于权威信息的展示，要求采集平台，定向采集。积累到 AQ 行业事件库。
美国 ABC 新闻网站大选板块。
[选举新闻、直播视频和更新 - ABC 新闻 (go.com)]

（2）初选投票同期对比



组件说明：
超出同期，显示人气比往期更旺盛。
投票和百分比的位置不准，应该在 110 和 244%的上方。

数据来源：要求采集平台，定向采集。积累到 AQ 行业事件库。
美国 ABC 新闻网站大选板块。
[选举新闻、直播视频和更新 - ABC 新闻 (go.com)]

（3）初选选举日历



数据来源：对于权威信息的展示，要求采集平台，定向采集。积累到 AQ 行业事件库。

美国 ABC 新闻网站大选板块。

[选举新闻、直播视频和更新 - ABC 新闻 (go.com)]

（4）认识候选人



点击候选人头像时，弹窗显示该人物画像。画像内容采用人物画像模板。

（5）候选人相关事件



组件说明

- 事件类型按照 Gdelt 的事件分类，分为 20 个大类。后续与客户讨论，是否有其他事件分类。
- 人物下拉，选择初选党派得票前 3 的人物。在总统竞选阶段，还支持选择选举团队，（团队包括副总统、发言人，助手等。）
- 事件下拉，支持选择对华事件、对俄事件、对朝事件等等，与中国接壤的 14 个地缘国家，全部事件。对华事件，是指描述中提到中国或 China，其他类似。

再次备注：所有下拉组件，均支持多选操作，方便用户对比！

5、全国党代会

（1）提名总统、副总统候选人



（2）竞选纲领



组件说明：

- 展示竞选口号：MAGA
- 展示纲领议题和主要内容，与中国以及中国国家安全观相关的，字体标红显示。比如经济，太空、极地等安全字眼

数据来源：对于权威信息的展示，要求采集平台，定向采集。积累到 AQ 行业事件库。

- 美国 ABC 新闻网站大选板块。选举新闻、直播视频和更新 - ABC 新闻 (go.com)
- 或者党派竞选网站。后续进行业务调研。

（3）竞选纲领历史对比

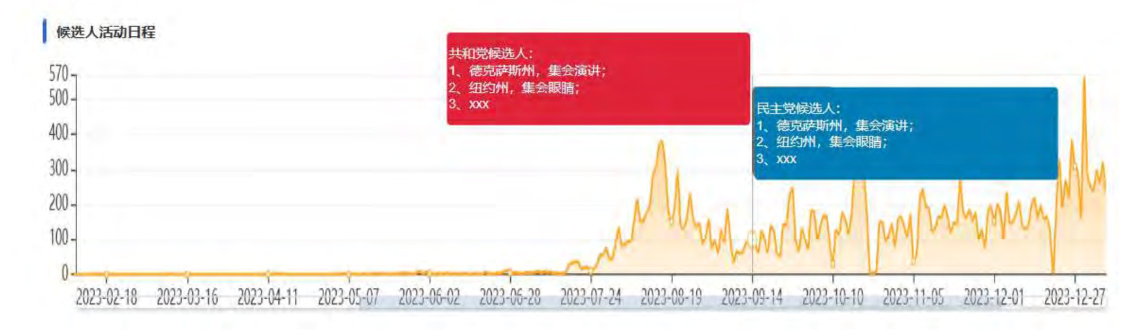


组件说明：与中国以及中国国家安全观相关的，字体标红显示。比如经济，太空、极地等安全字眼

数据来源：要求采集平台，定向采集党派竞选网站。若采集不到，则人工梳理，再导入数据库。实施状态显示议题在胜选后，是否实施。不确定的则用问号显示。

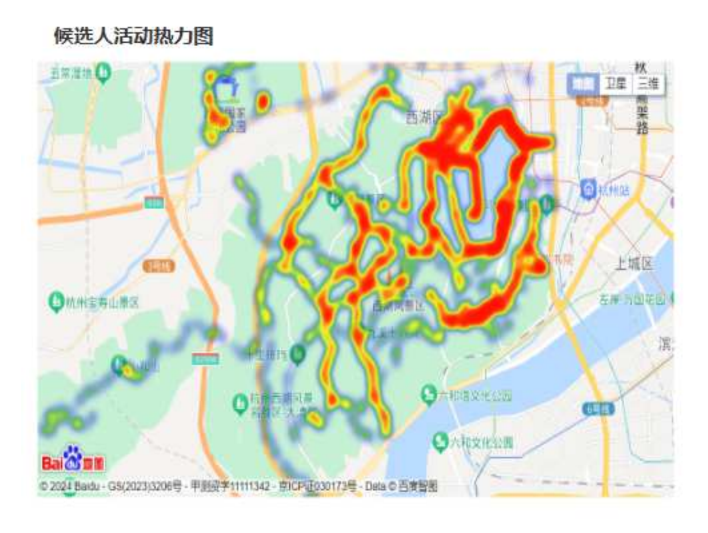
6、总统竞选

（1）候选人活动日程



纵轴是活动数量，是两党候选人之和。鼠标悬停时，左侧显示共和党候选人当天日程。右侧是民主党。

（2）候选人活动热力图



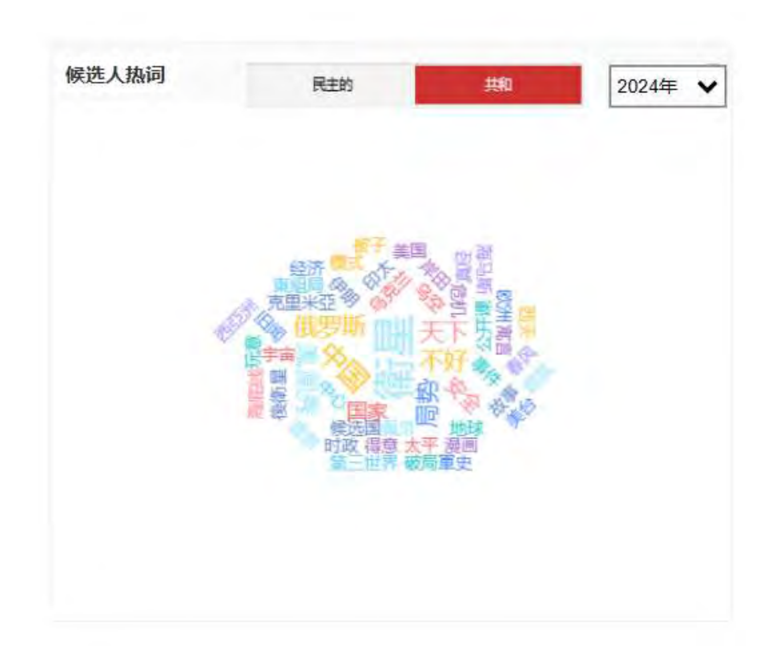
（3）选情预测



数据来源：要求采集平台，定向采集民调网站、标普数据、博彩网站。积累到AQ 行业事件库。

- 具体网站地址，后续进行业务调研。
- 对于历史预测，因为没有历史信息积累，需定向采集某些分析网站。具体网站，待后续调研。

（4）候选人热词



对于历史热词，因为没有历史信息积累，需定向采集某些分析网站。具体网站，待后续调研。

（5）大选话题 top10



- 数据来源：对监测词筛选后的信息列表，进行话题算法抽取，并统计数量。
- 对于历史数据，如果采集不到的话，需人工梳理后再导入。

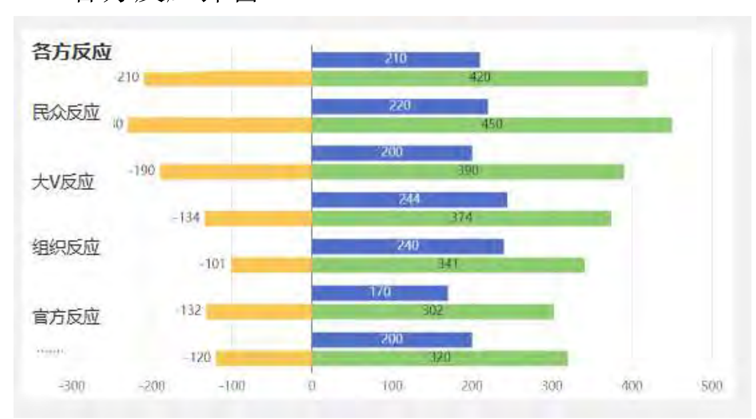
（6）Top10 话题溯源分析



组件架构	功能描述	备注
下拉框	选择话题，默认显示 top1	单选
各方反应		点击按钮，弹窗
节点	包括时间、信源名称、国家、标题。 点击标题，弹窗进入信息分析页面	节点数过多时，支持左右滑动。默认显示前 5 个。

备注：

- 节点支持导出、导入，方便用户修改抽取的信息，形成真实溯源展现。
- 报道或转发的媒体，大 V 可能数量很多，组件要求上榜的条件是阅读量超过 10000。
- 各方反应弹窗



左侧橙色是负面统计，右侧绿色是正面统计，右侧蓝色是中性统计。
官方定义：是指政府部分机构，比如外交部等。来自采集大库中的官方库。
组织定义：政府组织和 NGO 组织。
大 V 定义：全网粉丝超过 100 万。
民众反应定义：对话题信息列表中，评论数据进行情感识别。

备注：需技术经理确认是否有相关数据库或者标签。如果采集平台没有，则需按照定义规则，从采集大库单独维护相关数据库，作为行业积累。

（7）竞选资金分析



数据来源：要求采集平台定向采集，美国联邦选举委员会官网：首页 |FEC 系列

组件说明：资料来源不再显示。显示实时统计信息。标题即组件名称，去掉~~图表2020年特朗普与拜登。~~

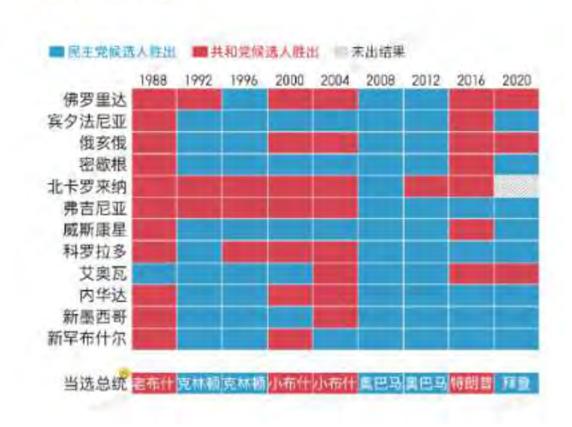
7、全民投票

（1）选举结果地图



- 数据来源：要求采集平台定向采集，美国 ABC 新闻网站大选板块。选举新闻、直播视频和更新 - ABC 新闻 (go.com)

（2）往年摇摆州分析



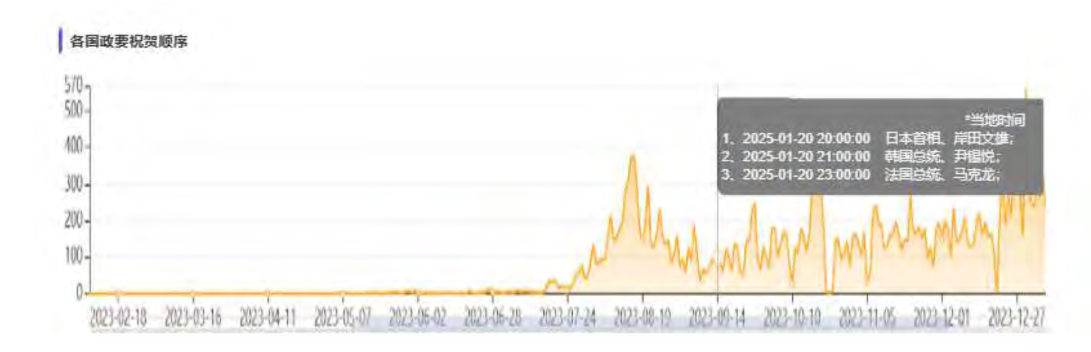
数据来源：对从 1988 年开始，对每个州历届选举结果进行对比，只要发生过变更的即视为摇摆州。
如果采集不到历届选举结果，可人工维护数据。

（3）支持者画像



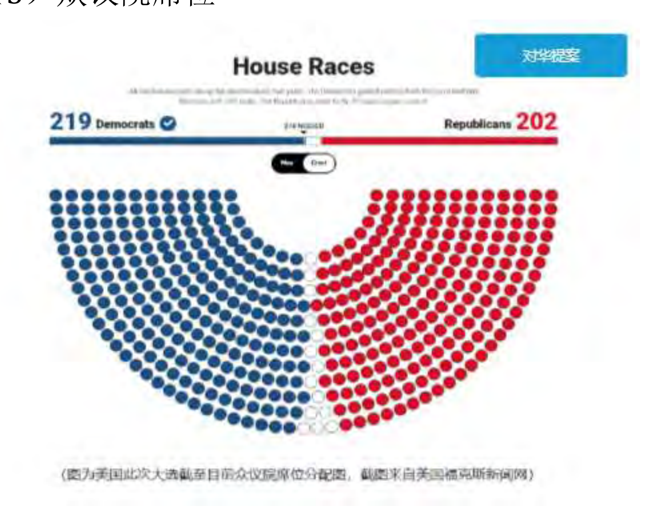
- 数据来源：要求采集平台定向采集，美国多家民调机构数据。具体网站，后续调研业务补充。

（4）各国政要祝贺顺序



注：起始时间是 2024 年 11 月 5 日。切换成美国东部当地时间。

（5）众议院席位



（图为美国此次大选截至目前众议院席位分配图，截图来自美国福克斯新闻网）

- 数据来源：要求采集平台定向采集，具体网站后续根据业务调研。

备注：美国众议院共计 435 个席位，所有席位统一每两年改选一次。参议院共计 100 个席位每两年选举其中 1/3

（6）参议院席位
同上。

8、总统就职典礼
（1）总统演讲摘要



（2）各国官方报道时间



3.2.1.7 舆情类模板组件

与天罗产品，保持一致。后续根据业务调研，持续补充。

3.2.1.8 组织类模板组件

无特殊组件，与人物组件类似。后续根据业务调研，持续补充。

3.2.2 模板管理

➢ 原型如下：



3.2.2.1 模板列表页

整体分为上下两部分。

- 上面是系统模板，只有管理员角色的账号，可以添加系统模板。
- 下面是个人自定义模板，管理员看不到普通用户的模板。
- 列表页支持对模板的增删改查功能。最顶部支持关键词搜索，搜索范围是模板名称和模板描述。

模板采用卡片展示形式。

模板组件	功能描述	备注
封面图	系统内置	自定义模板的封面都一样，系统模板看 UI 图。管理员新增系统模板时，随机选择内置的封面图
模板名称		不超过 20 个字
模板描述		
预览图标	点击后进入预览查看	鼠标悬停时，显示图标
编辑图标	点击后进入编辑	
删除图标		按照统一规范，删除需二次确认。

包括封面图（系统内置），模板名称、模板描述。

3.2.2.2 模板详情页

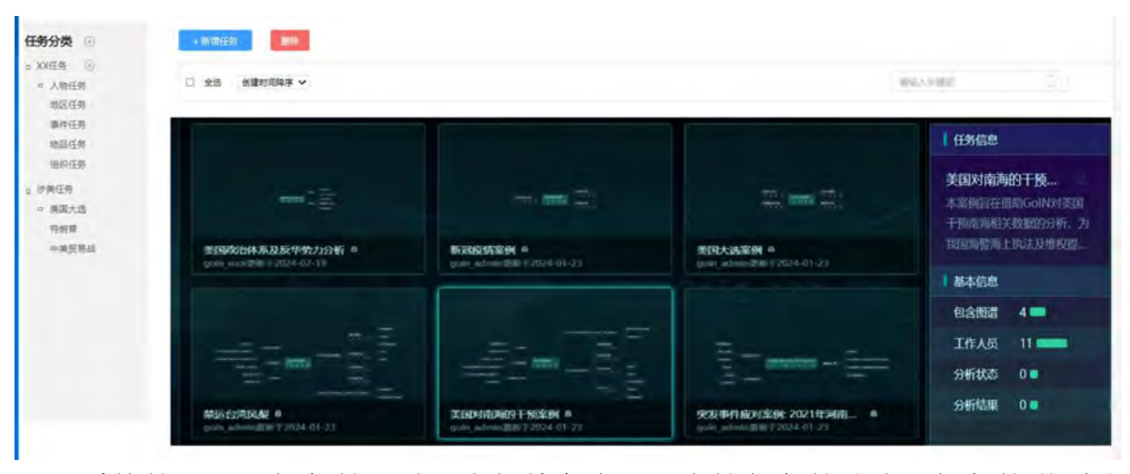
鼠标悬浮到模板卡片时，自动显示相关图标按钮，点击预览按钮，可查看模板详情。点击编辑按钮，可修改模板。个人只能修改个人创建的模板，无法修改公共模板。



3.3QB 服务

QB 服务包括任务管理、图谱管理和 QB 分析。QB 分析支持关联分析、时空分析。

3.3.1 任务管理



同一系统的设计逻辑保持一致，类似特色专题，支持任务的分类。任务的增删改查。

- 支持全选删除功能。
- 支持按创建时间排序
- 支持按关键词搜索，搜索范围是任务名称和描述。
- 选中单一任务后，右侧支持显示任务信息和基本信息。任务信息包括任务标题和描述。此处支持修改。基本信息包括工作人员，分析结果和分析截图。

3.3.2 图谱管理

需求未定，用户希望操作步骤极简。边建设，边沟通。

3.3.3 QB 分析

3.3.3.1 关联分析

需求未定，用户希望操作步骤极简。边建设，边沟通。

3.3.3.2 时空分析

需求未定，用户希望操作步骤极简。边建设，边沟通。

3.4报告管理

- 支持对报告的查询、搜索、预览、下载等功能。



- 支持报告的预览功能



- 支持报告编辑功能，支持对组件内容的编辑。



- 支持下载报告

3.5采集概览



完全复用采集平台的网络数据底图功能。直接内嵌相关功能模块，无须二次开发。

3.6系统管理

采用公司统一控制平台，包括机构管理，用户管理，角色管理。

4 软硬件环境需求

4.1软件环境需求

需求名称	需求定义
服务器端操作系统	CentOS Linux release 7.6.1810
服务器端数据库	关系型、非关系型
客户端浏览器	Chrome 浏览器、IE 浏览器 9 以上、360 浏览器等

4.2硬件环境需求

确定在不同并发量同时执行任务的最小配置。

需求名称	50-80 人并发
CPU	24 核
内存	管理节点：32G 计算节点：128G
硬盘	管理节点：500T 计算节点：2T

5 非功能需求

5.1信息采集

客户需求未完全确定，重点采集网站待后续梳理。下同。
需要杨晓菊老师全部的重点人物库、组织库、机构库，全部的业务标签体系。

序号	性能项	需求描述
1	采集信源	
2	采集规模	
3	采集频率	
4	回溯采集	

5.2文本处理

序号	性能项	需求描述
1	情感判断	支持正、中、负情感判断，准确率不低于 85%
2	文本分类	支持热点分类，准确率不低于 80%

3	文本聚类	支持热点聚类，准确率不低于 90%，聚合计算时间不高于 15s
4	主题抽取	支持对内容主题的聚合和抽取，准确率不低于 80%
5	内容去重	支持基于相似度计算的内容去重，准确率不低于 90%，去重计算时间不高于 15s
6	实体抽取	支持人名、地名、机构名等实体抽取，准确率不低于 90%
7	内容抽取	支持对时间、稿源、标题、征文
		后续还会持续补充。

5.3信息分析

序号	性能项	需求描述
1	实时分析	支持实时分析，可视化展现

5.4数据存储

序号	性能项	需求描述
1	存储周期	正式服务用户专题数据永久存储，其它数据存储周期 2 年以上

5.5响应时间

序号	性能项	需求描述
1	启动功能响应时间	低于 2s
2	数据保存响应时间	低于 5s
3	任务提交响应时间	低于 5s
4	简单查询响应时间	低于 5s
5	复杂查询响应时间	低于 10s

5.6系统并发

序号	性能项	需求描述

1	并发用户	同时满足 150 人在线用户请求

5.7系统安全

序号	性能项	需求描述
1	备份	系统应定期备份
2	日志	应用软件运行时须有完整的日志记录，以便及时审查日志文件，分析告警信息，掌握运行状况，对系统可能发生的故障做好应急方案。
3	权限	用户管理、权限管理应充分利用操作系统和数据库的安全性。提供统一有效的安全措施来进行权限限制，防止内部人员误操作或越权，每一个环节必须进行安全性检查，不能越权访问。对每次用户登录要求进行日志记录，对三次（参数可以改变）以上的登录尝试进行锁定（锁定的时间、解锁方式，参数可调），并发出告警。
4	开发与测试	开发、测试系统与生产系统要严格分开。
5	软件修改、升级	软件程序的修改或增加功能时，须提出修改理由、方案、实施时间，报上级主管部门批准；程序修改后，须在测试系统上进行调试，确认无误经批准后方可投入生产应用；软件修改、升级前后的程序版本须存档备查，软件修改、升级时须有应急补救方案
6	检验纠错	数据传输、处理应有检验、核对功能和较强的纠错功能。
7	监督告警	具有安全监督功能、故障和异常告警功能。

5.8数据安全

序号	性能项	需求描述
1	接口安全	系统采用三层（多层）体系结构，所有的数据操作都通过中间层来处理，客户端不能直接与数据库相连。外围系统也不能直接存取本系统的数据，必须通过系统提供的接口服务，来进行信息的交换。提供完整的系统数据传送监控和报警处理。
2	数据传送	关键数据在传输过程中，必须加密传送，以保证信息的安全。

3	数据备份和恢复	可提供各种方式、各种介质的数据备份功能。各种原始业务数据均应保留备份、各种过程业务数据应保留备份、操作系统、数据库应定期备份。提供对一般性数据和交易性数据的保护，使得在任何一部分子交易失败后，整个交易的所有消息都回滚，保证数据的完整性。 系统配置型数据，按月全量备份更新，按周增量备份； 用户配置型数据，按周全量备份更新，按天增量备份； 内容数据，支持冷热数据备份，超过 1 年数据即回滚成冷数据；
4	数据范围	登陆时检查 IP 地址的合法性，同时记录 IP 地址的使用情况。

5.9升级与扩展

序号	性能项	需求描述
1	可扩展	设计具有技术先进性和前瞻性,充分考虑系统升级、可拓展等问题，保证软件新旧版本的平稳过渡。
		保证主机系统，主机系统、网络系统、数据库系统、操作系统、业务应用系统和外围设备在将来能够顺利升级或扩容，且不影响正常的系统运行。
		用模块化设计,在整个系统正常运行的情况下,能够随时增加或取消接入用户。
		新的软件模块即插即用，新功能、新业务的增加能够在不影响系统运行的情况下实现
		可动态增加服务器，增加处理能力，能够实现负载平衡
2	版本管理	为了适应不同用户的不同处理流程的需要，系统应该支持使用版本组控制不同地方使用不同的版本组，在同一版本组内使用最新的版本，保证版本的自动更新。
		局部功能新旧版本的使用、并存、排斥要可控。局部功能新旧版本的使用、并存、排斥要可控。

5.10可维护

序号	性能项	需求描述
1	可维护	高内聚、低耦合设计，增强代码可修改性。
		确保软硬件的可移植性。

5.11可兼容

序号	性能项	需求描述
1	可兼容	系统所有交互均基于 Web 方式完成，所有页面跟功能需兼容 IE10 以上、Chrome 、Firefox 等主流浏览器

6 附录